www.environment-agency.gov.uk

Environmental Genomics - An Introduction

ENVIRONMENT AGENCY

Environmental Genomics - An Introduction

Research Contractors:
Cardiff School of Biosciences
Astra Zeneca
Centre for Hydrology and Ecology, University of Exeter

Commissioning Organisation:
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury
BRISTOL BS32 4UD.
Tel: 01454 624400 Fax: 01454 624409 Website: www.environment-agency.gov.uk

Dissemination status
Internal: Released to Regions
External: Released to Public Domain

Statement of use
This report provides an introduction to environmental genomics; describing how genomic tools and knowledge can be applied in the context of environmental protection. A review of current techniques and case studies of current exploitation of these tools is provided. This documents is for background information and reference for Environment Agency staff and others requiring information on this science and its potential applications to inform strategic planning.

Keywords
Environment, Genomics, Ecotoxicology, Risk Assessment.

Research contractors
This document was produced by:
Jason Snape & Dan Pickford
AstraZeneca,
Brixham Environmental Laboratory
Freshwater Quarry,
Brixham,
Devon, TQ5 8BA.

Peter Kille & Graham Small
Pest Management & Ecotoxicology,
Cardiff School of Biosciences,
Biomedical Sciences Building,
Museum Avenue,
PO Box 911, Cardiff, CF10 3US.

David Spurgeon
Centre for Ecology and Hydrology,
Monkswood,
Abbots Ripton,
Huntingdon,
Cambridgeshire,
PE28 2LS.

Charles Tyler,
Environmental & Molecular Fish Biology
School of Biosciences,
Hatherly Laboratories,
University of Exeter,
Prince of Wales Road,
Exeter, EX4 4PS.

Environment Agency's Project Manager
Dr Danielle Ashton, National Centre for Ecotoxicolgy and Hazardous Substances.

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre, c/o
WRc, Frankland Road, Swindon, Wilts SN5 8YF

Tel: 01793-865000 Fax: 01793-514562 E-Mail: publications@wrcplc.co.uk

ACKNOWLEDGEMENT AND FOREWORD

I would like to extend my sincere thanks to the team of authors who have been responsible for compiling this comprehensive report. In addition, I am grateful to the academic research teams, SMEs and multinational industries who have allowed their research and technical data to be incorporated into this report. It is my opinion that there should be no conflict between promoting the competitiveness of UK plc, whether in the academic research sphere, the nurturing of new start-ups or supporting our backbone of large industries, whilst ensuring sustainable protection of our local and global environment. However, this synergy can only occur through a partnership between these groups. Therefore, in advance of the major impact that "genomics" will have of this sector, my major recommendation is that we establish a forum that will bring together our legislators with the internationally renowned academic research community, small enterprises involved in environmental assessment and waste treatment and major industrial concerns.

Dr Peter Kille

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

The genome describes the full set of genetic "instructions" retained by an individual organism. The human genome therefore includes all of the genes that code for proteins, together with the control sequences for each gene, as well as the 'junk' DNA that occurs between and within genes.

Genomics is a broadly used term encompassing numerous scientific disciplines and technologies. These disciplines include: genome sequencing; assigning function to identified genes; determining genome architecture; studying gene expression at the transcription level (transcriptomics); studying protein expression at the proteome level (proteomics); and investigating metabolite flux (metabolomics).

Exposure to environmental toxins represents a stimulus that can induce changes in gene expression, which may be typical of that type of toxin. These changes can be monitored using genomic approaches, providing vast numbers of potential biomarkers. As the function of more genes is discovered, toxicogenomics offers the ability to unravel the mechanism underlying toxic effects of chemicals.

Inevitably, the science of environmental (ecological) risk assessment will need to develop in order to be able to take this new knowledge into account in a pragmatic way.

In this report the potential applications of the science within the Agency are outlined to attempt to indicate where it could have the greatest impact to the business in the medium-term. The caution that should be applied to this emerging science is also highlighted to help put the availability of the tools into context. A list of actions that need to be undertaken by the organisation is presented which will help it to move forward in a considered way and develop its role and agenda while the science continues to develop.

The issue

The science of the "omic" technologies is moving fast, a pace that will only increase with the level of funding that is being allocated in the UK and Internationally. The Environment Agency and other potential end-users of this technology need to determine:

- What the science could mean to the Agency in a regulatory context, clarity is required on the regulatory position i.e. how do we interpret such data if we receive it from a third party?

- Where can the Agency most effectively take up and implement the technology i.e. monitoring and diagnostic tools?

- How can the Agency develop its agenda on the subject in collaboration with researchers, funders and other regulators?

The scientific benefits and potential of genomics to increase our understanding of the biology of organisms is immense. It is essential, therefore, that the impact of developments in genomics on the Agency's business is evaluated and the time frame that this impact will be seen in the business is established.

The greatest benefit to be derived from genomics is information provision, genomics is capable of providing:

- A better understanding of ***mechanisms of action*** – enhancing the ability to identify the specific molecular targets of stressors, helping the Agency to understand the effects of exposure to multiple stressors and stressors at low levels.

- A more robust ***extrapolation of laboratory data to natural populations.***

- ***Predictive toxicity*** – the opportunity to measure sensitive sub-lethal molecular parameters will enhance the "preventative" mode of environmental protection and diminish the necessity for "cure".

- Through the identification of the molecular targets of pollution it may be possible to increase our insight into potential impacts on ***sensitive sub-populations and species***.

- ***Understanding the sustainability of populations*** by, for example, providing data on gene flow.

- Potential to ***reduce uncertainties in risk assessment*** and facilitate a more rapid evaluation of a chemical's toxic potential.

It is essential that the Agency develops its own strategy for genomics so that it can be proactive and take advantage of the opportunities presented. In addition, by being proactive the Agency will be in a position to protect itself against the potentially controversial issues that may arise through the use of genomics in an environmental context.

Specific actions recommended to the Agency include:

- To become more aware of the potential of genomic tools.

- To identify where the technology can be employed by the business and begin the necessary preparations which will enable its uptake.

- Collaborate nationally and internationally with genomic researchers to ensure the issues faced by regulators are an important aspect of the science agenda.

- Increase dialogue with other regulatory organizations.

1. WHAT IS GENOMICS?

1.1 Introduction

The term genomics is used to encompass everything from genome sequencing, annotation of function to genes, and genome architecture to studying patterns of gene expression (transcriptomics), protein expression (proteomics) and metabolite flux (metabolomics; Figure 1.1). For the purpose of this document the collective term "omics" will be used to address the transcriptome, proteome and genome.

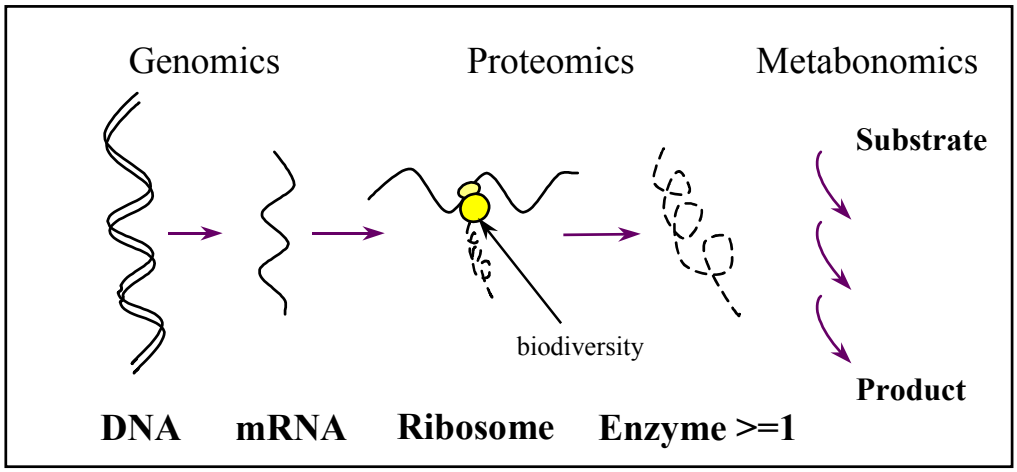

Figure 1.1 The molecular basis of life

Application of "omics" technology to (eco)toxicology

"Omic" technologies have the potential to reduce uncertainties in risk assessment and facilitate a more rapid evaluation of a chemical's toxic potential and the response of populations to environmental change. This learning will come from the incredible amount of molecular level information obtained from these technologies. This information will be used to elucidate new biological signalling pathways; new biomolecules; to understand mechanism of action; and the identification of sensitive and insensitive phenotypes and species. Given the amount of work required and the potential to misinterpret initial studies by equating a molecular change with an adverse effect, it was clear that the technical challenges ahead must be addressed through collaborations among scientists from industry, government, and academia.

Molecular genetics in epidemiology

Advances in molecular biology and in technologies for measuring and processing data at the molecular level will likely impact the use of biomarkers in epidemiology studies. Molecular epidemiology studies are proliferating in the scientific literature, exploring markers of genetic

damage, genetic risk factors for disease, and possible interactions among genetic factors, disease, and environmental exposures. Whilst this is being driven for epidemiological studies for human populations the approaches and issues are equally important for studying wildlife populations (eco-epidemiology). Multidisciplinary research teams will be needed to develop tools for, and interpret the data from, population-based molecular epidemiological studies. The application of genomics to environmental science will be reliant on continued high quality ecological and chemical monitoring programmes.

Ethical, legal and regulatory challenges in applying genomic technologies to toxicology and risk assessment

Clearly there will be ethical, legal and regulatory implications for these new technologies. Some of them are already being felt in our court systems and the problems will increase as we are able to generate customised gene expression patterns or individualised "fingerprints" detailing exposure to toxic chemicals, disease status, and potential susceptibilities. Areas discussed included the privacy of genetic information, protection of patient confidentiality, implications for regulatory agencies, applications in tort litigation, and discriminatory uses of genetic information by employers and insurers. The importance of the regulatory and legal challenges faced in the environmental application of these technologies also needs careful consideration. Simple issues such as regulatory bodies being able to identify 'polluting industries' that cause ecological impacts via specific conserved modes of action may be a future possibility.

Environmental Perspective

The Human Genome Project is the pinnacle of a new era of knowledge in medicine and biology. Fundamental research to understand the genome (the complete 'DNA blueprint' of an animal or plant) is also intensifying for many other 'model' species including fish, insects and worms. Inevitably, the science of environmental (ecological) risk assessment will need to develop in order to be able to take this new knowledge into account in a pragmatic way. This has significant implications for the Environment Agency in the future, with potential opportunities to select ecotoxicity test organisms that are best suited to a particular environmental question, save on the use of laboratory animals and drive forward economic benefits. Ecotoxicogenomics is the discipline that will act as the focus of this new knowledge.

Ecotoxicogenomics - origins

Toxicogenomics is the developing area of overlap between toxicology and genomic sciences. The genome is the complete set of gene sequences in the cell, the 'DNA blueprint'. As technological advances have allowed the sequencing (deciphering) of complete genomes of a number of organisms (including the human), science is now said to be in the 'post-genomic' era. What this means is that we now have the ability to monitor the expression of vast numbers of genes in parallel, using DNA micro-arrays, a tool to allow simultaneous analysis of many genes (see Section 2.3). These 'gene-chips' provide information on the transcription of genes and consequently this approach is also called 'transcriptomics'. All living organisms respond to even subtle changes in their environment through changes in expression (transcription) of many genes, for example those involved in homeostasis.

Exposure to environmental toxins also represents a stimulus that can induce changes in gene expression, which may be typical of that type of toxin. These changes can be monitored using genomic approaches, specifically transcriptomics, which effectively provides vast numbers of potential biomarkers. As we discover the function of more and more genes, toxicogenomics offers greater ability to unravel the mechanism underlying toxic effects of man-made chemicals. Rather than cataloguing which chemicals are toxic, and their potency, toxicogenomics has the potential to improve our understanding of how chemicals are toxic. As an extension of these principles, we can also use genomic approaches in ecotoxicology; hence the term 'ecotoxicogenomics' is born. Ecological risk assessment involves extrapolation of ecotoxicity data for a very limited number of test species to the protection of vast numbers of wildlife species. Through improving our mechanistic understanding, ecotoxicogenomics should enable more robust extrapolation of laboratory data to natural populations, through uncovering conserved mechanisms of action.

Challenges Ahead

Toxicogenomics and ecotoxicogenomics alike face many challenges before data generated by this type of approach is sufficiently robust to be used in a regulatory context. As already noted, changes in gene expression are highly 'state dependent': levels of expression may be significantly different between species, strains and individuals, between sexes and through ageing, through circadian and seasonal rhythms and in response to a wide variety of environmental stimuli. This means that DNA micro-array experiments are particularly vulnerable to the confounding effects of experimental error. Experimental models must be well validated, exposure systems tightly controlled, information concerning experimental conditions needs to be rigorously recorded, and normalisation and statistical analysis of data needs to be standardised.

These concerns may be magnified in the field of ecotoxicogenomics, where there are a greater variety of test species and, to date, no well established model species that have been sequenced and for which standardised 'gene-chips' are available. Furthermore, linkage between transcriptome data and more traditional toxicological endpoints is necessary in order to interpret whether changes in gene expression underlie an adverse effect at the level of the organism or population (with which ecological risk assessment is concerned). DNA micro-array experiments can provide vast amounts of data that are difficult and time-consuming to analyse. These issues need to be addressed if generation of such data is not to result in an increase in uncertainty in ecological risk assessment in the short-to-medium term.

Nevertheless, ecotoxicogenomics offers improved understanding of mechanism of action, greater predictivity and the ability to apply bio-informatics to safety assessment, identification of sensitive sub-populations, improved inter-species extrapolation, identification and validation of novel targets and biomarkers. For example, description and validation of transcriptome profiles for a range of reference toxicants of known mechanism of action could aid hazard assessment by providing an increased number of 'molecular biomarkers' (Figure 1.2). Critical steps in progressing this field include identification of gene loci affecting ecological performance, functional characterization of those gene products, assessing the extent and significance of intra- and inter- species variation of those traits to enable assessment of individual, population, community and ecosystem responses to the environment. A number of UK and international research programmes are addressing these issues (see Chapter 4).

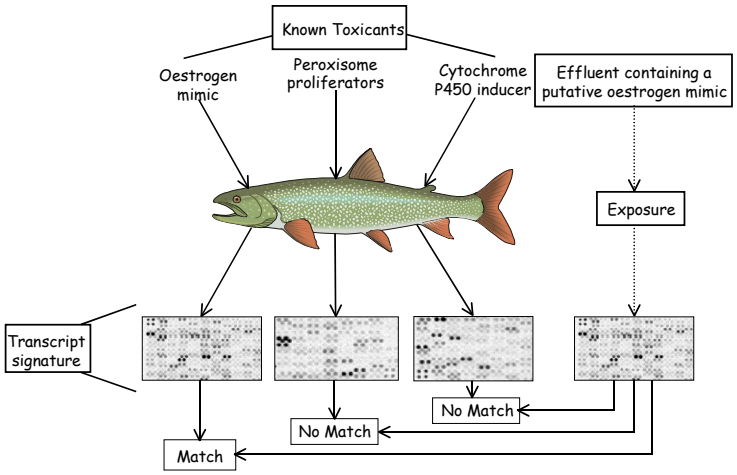

Figure 1.2 Simplified example of potential use of transcriptome profiles in ecotoxicological hazard characterisation

1.2 The vocabulary of genomics

1.2.1 "omics"

Although the suffix "omics" seems in recent years to have become overused, being employed to provide classical disciplines with a "popular" or "trendy" feel, its use on the whole is grammatically justified. "Ome" is derived from the Latin "mass or great number" and it was not until recently that biology had both the enabling technology and knowledge base to move away from considering individual biological components and to start to consider a more global analysis of a "great number" of end points. Most of the genomic "slang" can be easily unravelled by simply employing this definition i.e. genomics is therefore the study of a "great number of genes" whilst transcriptomics is the study of a "great number of transcripts".

The complementary information required when handling this myriad of new terms is a "road map" to the biological parameters commonly being investigated. A synopsis is provided here in the form of a schematic diagram (Figure 1.1) outlining the various levels of biological organisation of their relationship to one another. However, this really only provides a simplistic definition since these and other terms which have come into common use have evolved in meaning. Below is an outline of the general context in which these terms are used.

1.2.2 Genomics

The genome describes the full set of genetic "instructions" retained by an individual organism. Alongside the sequences that encode the building blocks of our cells "proteins" there are also instructions that control the expression of each gene in response to

environmental change as well as significant stretches of redundant information. A strict definition of genomics would describe studies relating to the structure of genomic DNA. This has led to the term becoming synonymous with the activities of the major genome sequencing centres which, as well as sequencing genomic DNA, are also active in identification and characterisation (sequence and functional) of the products of these genomes. Today, therefore, genomics is a broadly used term encompassing 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 (transcriptomics); studying protein expression at the proteome level (proteomics); and investigating metabolite flux (metabolomics). Due to the magnitude and complexity of '-omic' data, these disciplines are underpinned by information technology support through bioinformatics.

1.2.3 Transcriptomics

When a cell senses changes in its environment, it responds by accessing different components of its genome. For the most part, this access comprises the expression of genes encoding instructions for the production of new cellular "building blocks" or proteins. These instructions are read from the genome using a process called transcription to create messenger RNA (mRNA) molecules known as transcripts. Transcriptomics is loosely defined as the systematic quantification of the levels of all or a large proportion of the transcripts expressed within a cell under particular environmental conditions.

1.2.4 Microarray / macrooarray

All "omics" approaches require the quantification of large numbers of biological parameters at one time. One method is to display the biological entities (probes) within a defined pattern or array. Arrays have been constructed displaying DNA, proteins and antibodies on a range of surfaces including membranes, glass slides and gels.

Most commonly a reference to a microarray describes the precision dispensing of biological entities (probes) onto the surface of a charged membrane. Either through exposure to temperature or by UV cross-linking the probes are fixed within a defined location on the membrane. The biological entity is thus displayed within a defined grid from which a signal can be measured and attributed to each of the probes arrayed. DNA macroarrays usually employ a positively charged membrane and will display up to 10,000 genetic elements within a 20cm x 10cm area. Each DNA element is placed on the membrane using a robot to ensure both the position and volume of DNA is kept constant. In order to quantify the amount of complementary DNA within a biological sample (targets), it must first be labelled with radioactivity. This is performed by copying the DNA in the presence of a radioactive nucleotide. The label target is then hybridised with the membrane and the amount of label bound quantified subsequently using autoradiography or phosphorimaging. However, only one target can be measured, therefore it is of paramount importance that each array constructed displays the same quantity of biological material.

Microarrays employ glass slides or silicon wafers to support the arrayed probes. Glass microarrays can display up to 30,000 elements on a 10cm x 1cm area. The reduction in scale increases the technical challenge behind the liquid handling and image analysis, however the

fundamental theory involved is the same as macroarraying. The majority of microarrays utilize a duel fluorescent approach where two distinct fluorescent molecules are incorporated in complementary DNA generated from two biological samples (targets). Both targets can be hybridised to the microarray simultaneously therefore allowing a relative expression value to be calculated for each probe on the microarray. Due to its comparative nature, this relative measurement is far more robust to the quantity of DNA displayed on the array.

Affymetrix have developed a technique where oligo nucleotides may be synthesised directly on the surface of silicone wafers. This process allows a phenomenal density of oligo nucleotide probes to be packed into a small area (1cm2 60,000 elements). Furthermore, this technology is so precise that it employs proprietary biotin labelling and laser mediated signal analysis to analyse each array independently. Although this technology provides phenomenal quality control and specificity it is extremely costly and it relies on substantial prior information concerning the genetic make-up of the organism in question.

1.2.5 Proteomics

The building blocks of our cells are "proteins" which are encoded within our genomes and made by making an RNA copy of the genome sequence, this is then processed in Eukaryotes, to generate mRNA, which is then in turn converted into protein by ribosomes in a process known as translation. However, proteins can be modified after synthesis or translation, in a process known as post translational modification, either through specific cleavage (proteolysis) or by the addition of biological molecules (sugars, phosphates etc). Therefore, although it is predicted that the human genome encodes only 40,000 transcripts, due to mRNA processing and posttranslational modification this may generate up to 10 times that number of discrete protein molecules. Proteomics is the global analysis of these cellular building blocks.

Classically proteins have been separated using size fractionation through an acrilimide gel. More complex mixtures can be separated by exploiting the charge and size of a protein. By using this two dimensional approach, extremely complex mixtures of proteins can be separated. People have claimed to have separated up to 2,000 protein constituents using this technique, however, this separation only reveals the size and charge of the protein and does not facilitate its identification or allow it to be linked to the gene by which it is encoded. Recent advances in mass spectrometry have allowed extremely accurate size measurements to be performed on proteins separated using 2D electorphoresis. Furthermore, by exposing proteins to an enzyme that cleaves at a defined site, a fingerprint of protein fragments is generated which is characteristic of the protein from which it is generated. Malditof mass spectrometry allows a protein mass fingerprint (PMF) to be generated and this can be used to interrogate a database of the predicted PMFs from known genes and a match can be used to identify the protein. However, this process relies on having the genetic information of the organism from which the protein profile is being generated. For organisms where there is limited or no genetic information, it is essential to identify the exact sequence of amino acids that make up a protein in order to match to previously characterised proteins or genes. New developments in mass spectrometry have developed a technique known as tandem mass spectrometry (MS-MS) which allows regions of protein sequence to be generated in a relatively automated manner. However, this process is highly costly and still a development tool.

With an estimate of over 400,000 proteins within the cell and an ability only to separate approximately 2,000 on a gel, at present all proteomics is based on focusing on a subproteome or selection of proteins. This selection can be generated by classical biochemical approaches, i.e. subcellular fractionation/centrifugation or through more directed approaches, i.e. exploiting affinity chromatography in order to enrich for a family of proteins. When considering exploiting a proteomic approach or evaluating proteomic data, this limitation should be considered fully.

1.2.6 Metabolomics / Metabonomics

Metabolites are classically defined as the small molecules or non-polymers that represent the substrates and products of the chemical reactions occurring within a cell. These small molecules range from well characterised compounds such as sugars, lipids, amino acids and nucleotides, to more novel structures. Metabolomics is the study of the levels and composition of these metabolites within a cell. There has been a divergence in the nomenclature and approach to this area of research.

The terms metabolomics and metabonomics have both been used within this area. Although absolute definitions or distinctions are still being discussed, it is widely accepted that metabolomics describes the levels of native metabolites within a cellular system under varying environmental conditions, whilst metabonomics explicitly describes a systemic response profile taken from a sample collected from a complex, multi-cellular organism.

In parallel with approaches in proteomics, many metabolomic techniques utilise methods that focus analysis within a discreet class of metabolites, therefore many researchers exploit a selective extraction procedure followed by gas or high pressure chromatography coupled to mass spectral analysis. These approaches provide profiles of sub or discreet metabolomes. In contrast, more global approaches have been exploited by using nuclear magnetic resonance to generate a fingerprint of the complete metabolomes. The advantage of the former is that it can be executed rapidly and with limited specialist apparatus whilst the facilities required for the latter are highly specialised and exist only within a handful of laboratories in the country. Furthermore, with appropriate standardisation, the former can be used to identify specific metabolites relatively straightforwardly whilst the latter requires specialist knowledge and further analysis in order to allow metabolite identification.

1.2.7 Bioinformatics

When considering “omic” approaches, two distinct data issues become apparent. Firstly, a mathematical problem presents itself, connected with the statistical evaluation of the data and the interpretation and refinement of the results. Coupled with this is the association of each experiment with the relevant experimental description and related information concerned with each biological element being examined, this associated information is known as metadata. To address these issues it is essential to amalgamate mathematical expertise, computational competence and biological knowledge. Bioinformatics is the subject area that fills this essential interface. It is an essential component of any “omic” project and is the tool that will convert data overload into a useable output.

1.2.8 Toxicogenomics and eco-toxicogenomics

These are subdisciplines 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 Lannaconne 2001; see also Rockett & Dix, 1999; Lovett 2000; Pennie et al., 2000; ECETOC 2001). The toxicogenomic approach presents important opportunities to improve understanding of the molecular mechanisms underlying toxic responses to environmental contaminants (Bradley & Theodokaris 2002; Moore 2002).

1.3 Genomics and the biological effect cascade

It is well recognised that biological techniques can be potentially of great value when applied to studies assessing the efficacy of current environmental protection policy (Crane and Byrns, 2002; Peakall, 1992). Biological methods can for example be used to monitor changes in environmental quality over space and time and, in certain cases, can help link cause to effect. However, to date, although many biological tools have been developed for monitoring the state of the environment, few of these have been routinely applied within environmental protection regimes.

When selecting and evaluating biological methods for use in environmental assessment it is important to have clear criteria from which to select the method most fit for the purpose. Hopkin (1993) established a simple set of such criteria that have recently been used to assess the suitability of potential tools for monitoring the status of contaminated soils (Spurgeon et al., 2002). These evaluation criteria are that each biological method should be:

Reproducible (repeatable) so that the assay produces similar responses at the same level of stress upon repetition.

Representative so that the assay can be used to study the impact of the potential environmental change at a range of locations.

Responsive so that a measurable response is exhibited within a realistic range of environmental change.

Robust so that the assay should not respond to changes not associated with the investigated factor.

Relevant so that the assay provides ecologically meaningful data or results that can be related, preferably in a mechanistic way, to ecosystem level effects.

Ease of Use so that it can be executed rapidly and without a minimum requirement for specialist expertise for execution of data interpretation.

Cost so that the costs of the assay reflect the potential benefit of the results recognising that specialist and routine deployment demand different financial models.

Although desirable, it is improbable that any single assay will meet all of these evaluation criteria. This is illustrated by considering the position of biological methods conducted at

different levels of organisation (molecular, individual, population, community, ecosystem) within the linked dose response cascade proposed by Kammenga (2000). This indicates that biological assessment parameters at different levels of organisation occupy a continuum between sensitivity and relevance (Figure 1.3). Within this cascade, those methods most sensitive to chemical dose, such as gene expression measurements and metabolic change, are also those for which the link to ecological relevance is largest. In contrast, the most ecologically relevant methods, such as ecosystem functional assays (decomposition, mineralisation, catabolic activity), are least sensitive. In the context of the genomic assays of transcriptomics, proteomics and metabolomics this means that the methods are at their most powerful for the sensitive (in relation to dose and time) detection of environmental change (i.e. the responses occur at an early point in the damage cascade). The converse weakness of the methods is, of course, that they are less relevant to key ecosystem criteria than are functional assays that measure gross change in the dynamics of ecosystems. This disadvantage can be overcome to a certain extent by studies that aim to link molecular changes with the characteristic ecology of the organisms or ecosystems of concern. Adopting this mechanistic approach can, in effect, alleviate concerns regarding the ecological significance of molecular indicator changes.

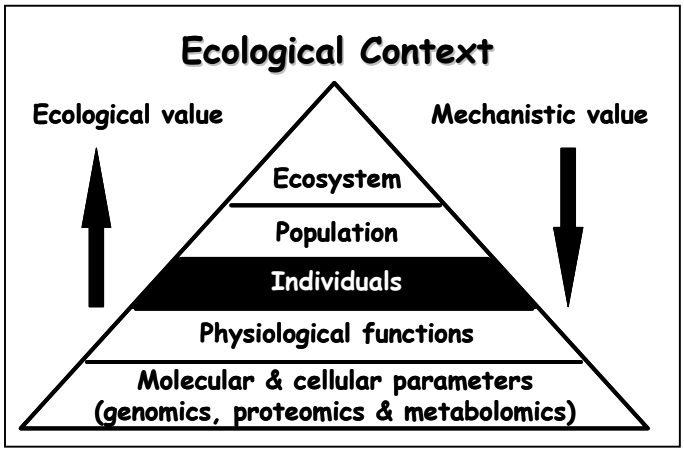

Figure 1.3 Biological response cascade

2. TECHNIQUES IN GENOMICS

2.1 Molecular techniques to study ecological condition

Even before the advent of genome sequencing projects and the subsequent development of transcriptomic, proteomic and metabolomic tools that exploit such information, the potential to assess the status of environmentally relevant organisms through measurement of their genes and proteins was already being investigated. The type of assay to which genomic approaches could be applied spanned the organisation cascade. Many of the methods and studies focused on assessing changes in the expression of single genes and also organelle level changes. Some, however, have exploited molecular methods to monitor change at the community level in both bacterial and eukaryotic species. The majority of these techniques require a considerable level of specialised knowledge (technical and theoretical), however, a number are being develop to provide solutions targeted at the monitoring market.

In the development of all of these molecular based assays, the most significant advances came upon the development of the polymerase chain reaction (PCR). PCR was devised by Mullis et al. (1986) and has proved itself to be the most versatile yet precise of all the biological techniques. During the PCR reaction (Figure 2.1), specific DNA sequences are multiplied by sequential splitting of the double stranded DNA molecule, annealing of specially designed complimentary oligonucleotide primers and extension to form a complimentary strand under the action of heat stable DNA polymerase. A series of heating and cooling cycles are used to drive the splitting, annealing and extension phases. To date PCR based methods have been applied in a number of techniques with potential for biological assessment of environmental quality. These applications are briefly reviewed below.

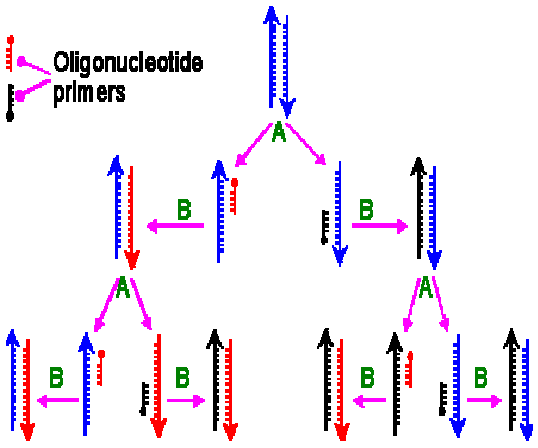

Figure 2.1 Schematic diagram of the PCR method use to amplify DNA fragments using sequence specific primers

Figure 2.2 Two PCR machine suitable for routine amplification of nucleic acids

2.1.1 Microbial community profiling

A number of techniques suitable for the analysis of the composition of the community are available and these have been widely used in many research projects. These can be either culture dependent or culture independent. It has been widely reported (Wagner et al., 1993) that culture dependent methods for studying microbial community structure only recovers approximately 10% of the total population and as such are inadequate for describing microbial community structure. This is often referred to as "the great plate count anomaly". Culture independent studies extract total DNA or RNA from microbial communities and use universal forward and reverse primers in combination with the PCR reaction to amplify species- or genera specific (usually the 16S subunit of ribosomal; Woese, 1987) DNA fragments from a whole community sample previously isolated directly from soil. After PCR, fragments can be used in a range of post amplification analyses (Figure 2.3; Stackebrandt & Goodfellow, 1991). Two quantitative "gold standard" techniques exist that are non-invasive i.e. they do not rely on the extract of DNA from microbial cells and thus can be used to examine microbial ecology in situ. These are flow cytometry and Fluorescence in situ hybridisation (FISH).

Denaturing and temperature gradient gel electrophoresis (DGGE and TGGE)
PCR products are separated by electrophoresis on an acrylamide gel containing a denaturing urea gradient (DGGE) or a temperature gradient (TGGE). The denaturing conditions induce strand melting at a point dependent on the nucleotide composition. These melted fragments migrate slower through the gel matrix, thus separating the fragments of differing nucleotide composition (Muyzer et al., 1993). A specific example of the use of DGGE to analysis populations of ammonia-oxidising bacteria is described below.

Single strand confirmation polymorphism (SSCP)
PCR is conducted with a phosphorylated and non-phosphorylated primer. Products are converted to single strands by lambda exonuclease digestion of the phosphorylated strand. These single strands are electrophoresed on a non-denaturing gel where they separate according to sequence-dependent folding confirmations that affect their mobility.

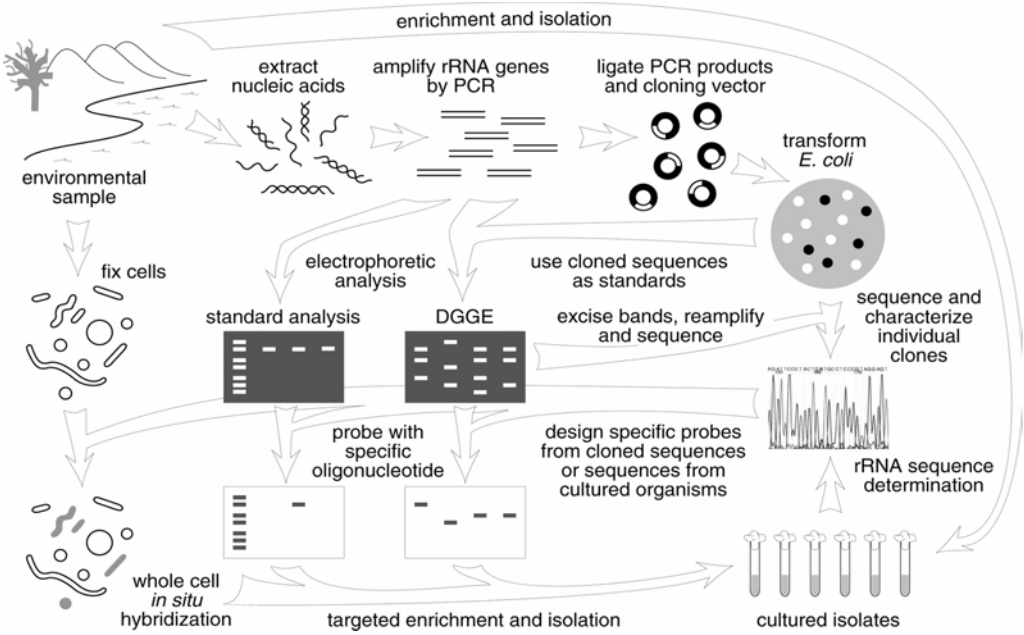

Figure 2.3 Schematic representation of approaches to microbial community profiling

Amplified ribosomal DNA restriction analysis (ARDRA)
PCR amplification is followed by cutting with restriction enzymes. Digests are then electrophoresed on agarose or acrylamide gels allowing identification of sequence dependent banding patterns. One methodological problem with ARDRA for quantitative estimation of microbial diversity is that the number of bands in the profile is always greater than the number of amplified PCR products. This technique is usually used to analyse clone libraries to screen for identical clones sets in order to prioritise and reduce redundancy in subsequent DNA sequencing efforts.

Terminal-restriction fragment length polymorphism (T-RFLP)
T-RFLP is similar to ARDRA, but a fluorescent primer is included in the PCR reagent mixture. After restriction enzyme digestion, fragments are analysed on an automated sequencer. Only those bands carrying the fluorescent label (i.e. the terminal restriction fragment) are detected. This number corresponds directly to the number of species.

Fluorescence in situ hybridisation (FISH)
FISH allows the direct scrutiny of microbial populations within their three-dimensional ecological niche. An environmental sample is fixed, using paraformaldehyde and the cell membranes permeabilised. Fluorescent oligonucleotide probes are introduced into the samples that are specific for certain genera or species of bacteria. Under certain conditions these probes will only hybridise with the specific bacteria against which they are targeted towards. After washing excess probe and filtration, samples can be visualised by either epifluorescence or confocal laser-scanning microscopy (CLSM; Figure 2.4). CLSM is particularly useful for

samples such as activated sludge where auto-fluorescence is a problem, as the laser will only excite the sample in a single focal plane.

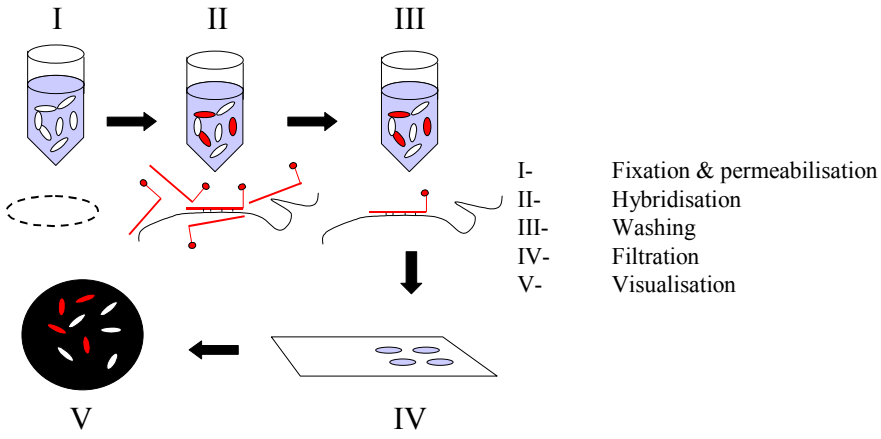

Figure 2.4 Fluorescence in situ hybridisation (FISH) of microbial populations

Flow cytometry

Flow cytometry is a generic technology which counts and measures multiple characteristics of individual particles in a flow stream. The initial application of flow cytometry in aquatic science was the analysis of phytoplankton, since the red chlorophyll fluorescence from chlorophyll can be used for the detection of photosynthetic cells even in the presence of large quantities of detritus. This is now being used as a tool to identify different taxonomic groups of bacteria and phytoplankton using spectral differences in auto fluorescence excitation and emission. Flow cytometry can also be coupled with the use of fluorescent oligonucleotide probes to detect and quantify specific species or genera of microorganisms in complex environmental mixtures. Flow cytometry is an inherently fast technique. Thousands of cells can be analysed per second. Routinely, 10,000 cells are analysed in one sample and up to 30 samples can be analysed per hour. Moreover, particles of interest can be physically sorted for subsequent analysis or purification of cultures.

2.1.2 Eukaryotic community profiling

Identification of the presence of eukaryotic species has traditionally relied on morphologically based taxonomy, a task that is often laborious. As an answer to this problem, it has been suggested that PCR based 'molecular barcoding' techniques could be used. The method is based on the amplification and sequencing of specific regions of the small subunit ribosomal (SSU) RNA gene. Identified sequences may then be categorised into molecular taxonomic units based on sequence similarity and bioinformatic search of known SSU sequences. The approach allows the presence of specific genera or species to be identified within a community and can be linked either to classical biodiversity statistics, specific indices (such as the nematode maturity index) or potentially to multivariate based analytical tools such as the RIVPACS system. To date the molecular barcoding has been developed for nematodes in

soil systems (Floyd et al., 2002). Other groups that may be suitable include oribatid mites and collembolans in soil and chironomids in freshwaters.

Genome mutation analysis using RAPD or AP-PCR

Exposure to genotoxins can result in covalent binding between DNA and the parent chemical and/or a metabolite to form adducts. Faulty repair of adducts can result in point sequence mutations. A number of techniques exist that can be used for the direct detection of DNA adducts, however developments in molecular biology now allow the detection of sequence changes resulting from genotoxicity using PCR. Both fingerprint methods use short (often 10 nucleotide) randomly sequenced primers. These anneal to DNA at a number of locations and produce a series of PCR products. Changes in sequence or DNA strand breaks will affect either PCR product amplification or size. These changes are detected by gel electrophoresis. To date, use of randomly amplified polymorphic DNA (RAPD) and arbitrarily primed PCR (AP-PCR) to detect genotoxins in the environment has been limited. RAPD has been used in a limited number of studies to detect genotoxicity (Atienzar et al., 1999, Conte et al., 1998, Theodorakis and Shugart, 1998). Savva (2000) meanwhile used AP-PCR to detect DNA damage in rats and shore crabs exposed to benzo[a]pyrene.

2.1.3 Single gene transcript quantification methods – selection of genes for analysis

Quantification of the expression of gene transcripts and translated proteins represents one of the most sensitive approaches for measuring the effects of environmental stress on organisms. Measurement of changes in the expression of potentially pollutant responsive genes can be undertaken for a vast range of different sequences and using a variety of protocols. To briefly summarise these possibilities, first potential target sequences will be set-out with examples, available measurement techniques will then be listed.

*Target sequences***.** Over the past 10-15 years, extensive research has been undertaken to identify genes differentially expressed in organisms maintained under different environmental conditions. Initially studies focused on human or rodent models. However, more recently the field of environmental genomics has emerged in which homologues of pollutant responsive sequences have been identified in environmentally relevant species, i.e. it is likely that the number of known sequences will greatly increase in coming years.

Among known environmental and chemical responsive sequences, three primary groups of genes that may be differentially expressed can be identified. These can be considered biomarkers of *chemical exposure, physiological compensation and effect*. Examples within each of these three categories (for environmentally relevant species where available) are given below.

Exposure. As this category includes genes encoding protein known to be involved in pollutant and detoxification pathways, a number of established biomarkers of chemical exposure are represented. For metals well known examples are the genes encoding metallothioneins (Kille et al., 1999; Maywald & Weigel, 1997; Stürzenbaum et al., 1998d), pytochelatins (di Toppi and Gabbrielli, 1999, Maywald and Weigel, 1997), ferritin (Kumar and Prasad, 1999, Richards et al., 2000), intracellular copper chaperones (Culotta et al., 1999), arsenic responsive genes from the ars operon (Sarkar, 1999) and the mercury responsive genes in the

mer operon (Kiyono et al., 2000). Genes in the pathways for detoxification of organic compounds include the cytochrome P450s (Fujita et al., 2001, Risso-de Faverney et al., 2000, Schlezinger and Stegeman, 2000) and glutathione-s-transferases (Gallagher and Sheehy, 2000, Hodgson and George, 1998, Pathiratne and George, 1998).

Physiological compensation. Subtle alteration in the physiology of an organism allows is to minimise the toxic effects of a compound by invoking compensatory systems. The molecular response underlying this response includes a range of genes involved in adaptive biochemical pathways. Best established are the heat shock family of protein chaperones involved in protein folding (Chen et al., 1999, Tully et al., 2000). Mitochondrial genes (Richards et al., 2000; Spurgeon et al., 2001a; Stürzenbaum, 1997), lysosomal function genes (Sturmbauer et al., 1999), metal containing enzymes (Stürzenbaum et al., 2001) and enzymes of the antioxidant system (e.g. super-oxide dismutase, catalase, peroxidase) also fall into this category.

Effect. Responses-linked biomarkers rather than being related to pollutant handling or compensatory systems are instead directly linked to observed physiological effects. The best known example is vitellogenin (VTG - see Section 3.3), particularly environments where exposure to environmental oestrogens is known or suspected (Celius et al., 2000, Denslow et al., 2001b). Other examples include zona pellucida proteins (Arukwe et al., 2001a, Arukwe et al., 2001b, Denslow et al., 2001a) and the oxytocin-related peptide annetocin. As more becomes known about the genetic control of growth, development and reproduction in non-model species, more sequences that fall into this category are likely to be identified.

2.1.4 Single gene transcript quantification methods

Sample preparation. Quantitation of expression of specific gene transcripts has been used successfully to provide environmentally relevant information (see case studies in Section 3). Advances in the handling and detection of nucleic acids has made a number of methods available for detection and quantification of gene transcription. In all protocols, the first step is to isolate total RNA using one of a range of commercial kits and reagents. Messenger RNA (mRNA) can then be separated using membrane or magnetic bead technology. This can then either be directly probed or used for RT-PCR after conversion to complementary DNA (cDNA) by reverse transcription. During all RNA handling steps great care must be taken to protect sample from degradation by ribonucleases (widespread RNA degrading enzymes). Typically RNA samples should be handled in a dedicated laboratory, using double autoclaved equipment and double autoclaved or certified RNAse free reagents. The range of potential detection methods for quantification of gene expression through mRNA measurement are detailed below.

Northern, blotting. Northern, dot and slot blotting are all hybridisation techniques. For Northern blotting, total RNA is electrophoretically resolved under denaturing condition and then transferred via capillary action to a nitrocellulose membrane. The membrane is probed with a radiolabelled oligonucleotide probe designed with a sequence that matches the target gene product. When a gene is highly expressed, more mRNA is present on the nitrocellulose and as a result there is greater hybridisation of the labelled probe. When viewed by exposure of audioradigraphic film to the membrane, such samples show a larger ‘blot’ than low expression samples. Image analysis can be used for formal quantification.

Dot and slot blotting. In dot and slot blotting, samples of cloned DNA matching the gene of interest are denatured and identical amounts uniformly spotted onto a single nitrocellulose membrane. The filter is then hybridised with a radioactivity labelled probe, containing the corresponding sequence in unknown amounts. The extent of hybridisation is estimated semi-quantitatively by visual comparison to similarly spotted radioactive standards.

RT-PCR for measurement of gene expression. The power and specificity (through the design of gene specific primers) of PCR makes the procedure ideal for detecting responses of specific gene. Initially an obstacle to the use of PCR for gene expression quantification was the nature of amplification. This is characterised by a logarithmic increase to a plateau (Kochanowski and Reischl, 1999). Because the plateau is derived from limitation not related to the quantity of template, this means that parallel reactions with vastly different template inoculations result in near identical final product levels. As a result, comparisons of product levels at amplification end would show no difference between samples containing different amount of starting material (Kille et al., 1999).

To overcome the problems of PCR for gene quantification, techniques have been developed that take a snap shot of product levels during reaction. Initially these quantitative PCR (Q-PCR) methods used either amplification for a limited number of cycles followed by gel electrophoresis and image analysis or quantitative competitive PCR (Evans et al., 2001). Such methods have now largely been supplanted by fluorescence in situ monitoring using specifically designed instruments. These platforms monitor PCR progress using two main fluorescence detection systems.

SYBR® Green dye. This fluoresces when bound inside the double helix of DNA. Fluorescence is proportional to the amount of DNA so repeated monitoring after completion of each PCR cycle allows product quantification.

The fluorogenic 5' nuclease assay. This utilises the 5′ nuclease activity inherent as a secondary function of Taq DNA polymerase. For the procedure an oligonucleotide probe complementary to the target is included within the reaction (Fig. 2.5). The 3′ end of the probe is labelled with a fluorescent molecule, such as FAM (6-carboxy-fluorescein), and the 5′ end with a complementary quencher, such as TAMRA (6-carboxy-tetramethyl-rhodamine). When both are present on the probe, the quencher suppress the fluorescent molecule and no light can be detected. In PCR, during primer anealling, the probe hybridises to the target between the two primer sites. On primer extension, the 5′ nuclease activity of the Taq DNA polymerase cleaves the probe liberating the fluor from the quencher. This results in detectable fluorescence. Inclusion of the TaqMan probe in the PCR reaction allows production of an amplicon to be monitored at each cycle.

For both detection systems, calibration standards containing cloned copies of the target gene at known concentrations can be used to obtain the relationship between transcript frequency and the number of cycles required to obtain a specific threshold. This standard curve is used to determine gene expression levels in the samples.

In Q-PCR normalisation of cDNA template concentrations are required to account for differences in sample handling, transcription efficiency and sample loading. For this, parallel measurement of a control (non-responsive) gene(s) can be employed. A number of invariant ‘housekeeping genes’, including actin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH),

ribosomal genes, cyclophilin and elongation factor 1-alpha, have been identified for this purpose (Stürzenbaum and Kille, 2001).

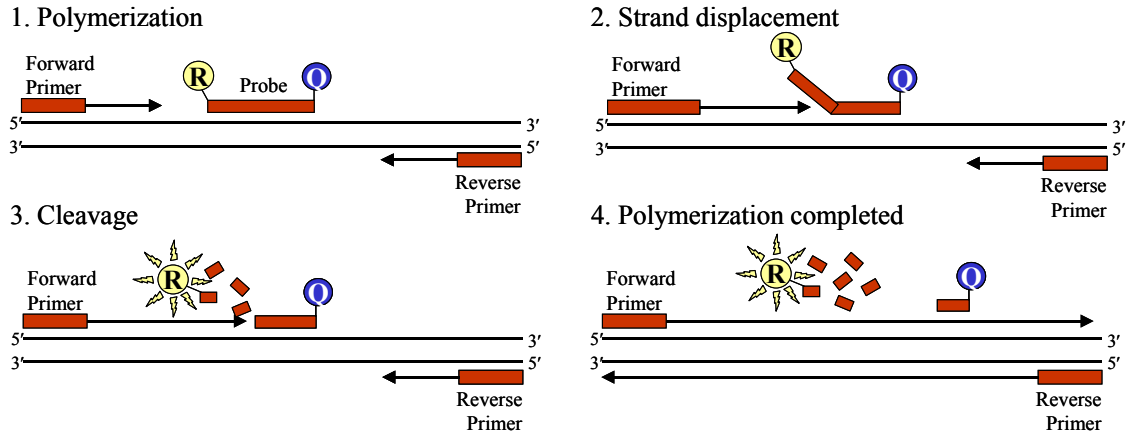

Figure 2.5 Schematic representation of mechanism underlying fluorescent detection of PCR product using Taqman™ hydrolysis probes (adapted from the ABI Prism™ 7700 Users Manual)

Direct quantification of mRNA transcripts by chemiluminescence. Chemiluminescent RNA hybridisation offer an alternative to quantitative RT-PCR for detection of specific mRNAs. The method is already applied in aquatic ecotoxicology for assessing VTG expression in fish exposed to xenoestrogens (see Section 3.3). For the assay, labelled probes complementary to target gene are added to an RNA sample. This hybridises to the target and a reaction occurs from which light is given off. The light generated is proportional to the number of gene transcripts present and thus gives a measure of the concentration of the gene product present in the sample.

2.1.5 The move to the "omic" technologies

The studies and methods outlined above are all focused on the measurement of transcription of specific genes whose function has been putatively linked to the environmental stress of concern. This targeted, reductionist approach, while excellent for specific hypothesis testing is, however, limited in the scope of environmental survey against which it can be targeted. For example, while it may be possible to evaluate if an organism has been exposed to a particular stressor (disease or chemical) of concern, it would be difficult to monitor a site for the presence of a complex range of potential stressors such as mixtures of chemicals with different modes of action, parasites and diseases and/or changes in land management practices. The desire to be able to detect changes in multiple gene endpoints within a single assay in order to investigate complex biological responses and disease states has led to the development of a series of exciting and promising global analytical approaches. These are capable of measuring changes in the full complement of expressed genes (in the transcriptome), proteins (in the proteome) and metabolites (in the metabolome). These methods, which are discussed in detail in the following sections, all allow the multivariate analysis of the biological status of organisms maintained under different environmental

conditions. These studies produce large volumes of data, which after suitable statistical evaluation, can tell us a great deal about the health of organisms living in a specific environment.

2.2 Are genomics technologies "fit for purpose"

The publicity surrounding the completion of the human genome has led to expectation that once the genetic sequence was revealed there would be miraculous advances in medical and life science research. It is now fully appreciated that determining the DNA sequence is the first and essential step in developing increased understanding of biology. In the same way, genomic approaches provide significant potential to advance both medical and environmental research, however, we must manage our expectation of what can be delivered by genomics.

Global profiling of communities, transcripts, proteins or metabolites is not compatible with the routine monitoring of the natural environment. A number of mitigating factors mainly genetic and environmental variability ensure that the 'normal' global profile alters significantly within the natural environment. Therefore linking changes observed in the global profile to harmful environmental impacts is unrealistic. However, genomics has a significant role to play in the laboratory studies in identifying the potential toxicological properties of compounds. Furthermore, it represents the most efficient method by which to identify new potential diagnostic markers which, once validated, may represent an output which is 'fit for purpose' for routine environmental monitoring.

Limitations

Genetic variation: Most genomic studies to date have exploited 'model organisms', the majority of which are inbred or clonal populations (e.g. array studies with the nematode *Caenorhabditis elegans* have generally employed the N2 strain which has been cultured in the laboratory for more than 40 years; Kim et al., 2001). This means there is little genetic difference or variability in these populations. However, this is not the case for the majority of organisms used within ecotoxicology and this variability increases when we consider natural populations. The consequence of this natural variation is already recognised within established ecotoxicology where toxicity profiles using natural populations exhibit substantially increased variation over those performed with laboratory populations. It is therefore essential to identify and describe the potential markers that form critical pathways which are mechanistically linked to specific biological processes and display limited variability. Genomic approaches are excellent vehicles for generating these potential markers. However, preliminary identification using genomics must be followed by substantial validation and characterisation of the natural variation before deployment.

Environmental variability: Subtle changes in the environment elicit substantial alterations in global profiles. This issue has caused substantial difficulties when analysing transcriptomic data where there have been a number of examples of environmental or experimental changes causing alterations in the transcript profile that are more significant than the changes in the designed experimental parameters. Factors ranging from individual animal handlers, time of day samples were taken and dissection practices have all been responsible for compromising extremely well designed and controlled toxicogenomic experiments. These mitigating factors can be controlled within laboratory settings and, to a large degree, GLP protocols established

for ecotoxicology go a long way to providing the correct regime for generating significant genomic data. However, direct environmental sampling and characterisation introduces a new dimension of variables, some of which cannot be controlled for. The genomics community have addressed this issue by introducing standard practices when recording the information or metadata associated with microarray experiments. The Minimum Information About a Microarray Experiment (MIAME) is now an internationally agreed standard relating to the information required to be recorded relating to a genomics experiment. It is already evident that environmental deployment of genomic technology will require additional levels of information to be recorded for each experiment.

Fit for your question

In order to directly link genomic approaches to the major activities of the environmental community we have reviewed the major work areas of the Agency and describe what we believe may be the relevant genomic approaches that would benefit these areas of investigation (Section 4).

2.3 Genomics tools that are at or near market

2.3.1 Global profiling tools

Substantial research tools exist for the global profiling of gene transcripts. A selection have been developed to a standard where their exploitation is quality controlled to a degree which allows them to be deployed consistently. The majority, however, are excellent research tools which enable profiling to be performed within bespoke research projects but, due to their intrinsic methodology or quality control variability, cannot be deployed with a high degree of reproducibility. The latter category includes (directed) differential display, subtractive library construction (suppression subtractive hybridisation amongst others) and serial analysis of gene expression (SAGE). These approaches employ novel comparative methods for displaying or enriching for genes or gene fragments which are expressed specifically within a test condition, they rely on the subsequent characterisation of the gene and validation of its involvement using single or limited transcript analysis (see section 2.3.2). Although some groups have deployed these techniques, to investigate temporal and dose responses. It is recognised that this procedure is not routine and generates very preliminary data. Proteomic and metabolomic profiling are also at present research tools which must be configured for each individual research project. In contrast the techniques described below can be deployed on multiple samples and given the right experimental design, compared between different investigating teams.

Macroarrays: Filter arrays are now commercially available for a range of bacteria and for a limited cohort of mammalian or model organism genes involved in specific pathways (usually those involved in stress and cell death). These tools are highly quality controlled and supplied with comprehensive standard operating procedures (SOPs) which ensure that reproducible data can be generated using GLP. Many commercial and academic laboratories can generate bespoke macroarrays, however, substantial quality control and validation of these materials is required before they are acceptable for routine deployment. The main commercial suppliers of these materials include sigma-genosys (www.sigma-genosys.com), MWG

(www.mwg.com), Compusys (www.compusys.co.uk) and Operon (www.operon.com), although many smaller industries are now appearing.

Glass slide Microarrays: Commercial suppliers now provide a substantial number of standard microarrays for model organisms, where the full genome has been completed, these represent all known transcripts generated by these organisms, these include a number of microbes, yeast, tuberculosis, malaria, *C. elegans* and *Drosophila*. Additionally, arrays representing the majority of transcripts from human, mouse, rat, *Arabidopsis*, zebrafish and *C. albicans* can be procured. A number of academic consortia supply many of the above arrays as well as developing microarrays for carp, rainbow trout, earthworms, *Daphnia magna* and medaka. As with macroarrays, the commercially available resource is supplied with substantial quality control guarantees and SOPs. The academic resources in contrast can be extremely variable in quality, some being superior to the commercially generated material and others being substantially compromised. The commercial suppliers are the same as for macro arrays whilst the academic groups involved can be found by looking through the relevant project summaries in Section 3.

2.3.2 Single or Limited Transcript analysis

The output of well designed genomic experiments should be a limited number of potential markers linked to specific biological pathways and specific environmental scenarios. The next step to converting this research output into a tool which is fit for purpose as a diagnostic test, is the validation of the target characterisation of natural variation and the development of a test platform. These processes will exploit technologies which focus on the measurement of a limited number or single genes. Within this section the technologies that are at market are reviewed, however the theoretical issues relating to internal controls, calibrators and quality control will not be explored. Classical methods for assessing transcript analysis such as Northern blotting or RNase protection are wholly inappropriate as they are qualitative techniques which cannot be translated into high throughput diagnostics. Techniques for measuring individual proteins using immunological methodologies are well established and will also not be covered.

Q-PCR: PCR is an enormously powerful tool for the specific logarithmic amplification of known sequences of DNA. Theoretically, amplification can be achieved from as little as a single starting template to greater than a billion copies within 30 cycles. However, due to substrate limitations and other factors, the amplification process plateaux independent of the quantity of starting material. If the amplification is monitored continuously or in real time, a linear relationship can be established between the log of the starting concentration and the point at which product reaches a critical threshold (within the logarithmic section of amplification). Using appropriate standards, this allows a calibration curve to be generated over 6 orders of magnitude of template. The limitations in this technique are consistent with any amplification technology and relate to the amplification of signal variation. This results in a limit to the accuracy of measurement that can be acquired using this technique, usually this is approximately 1.5 to 2 fold. A multitude of chemistries and platforms have been developed to allow real time monitoring of the PCR reaction. The chemistries include; hydrolysis probes (TAQMAN), hybridisation probes, hicks probes and the non-specific incorporation of a fluorescent molecule specific to double stranded DNA (Cybergreen). All of these approaches are compatible with the major Q-PCR platforms that are available, these divide into two major types, capillary PCR and tube PCR. The capillary PCR platform, trade name Lightcycler

(www.biochem.roche.com/lightcycler), performs the amplification cycle within a thin glass capillary which can be rapidly cooled and heated whilst fluorescence measurements are continually collected. This enables a 40 cycle amplification reaction to be completed in 20-30 minutes, however, since each sample must be individually introduced into the capillary, this reduces the ability to alternate the system. However, the originators of this technology, Idaho Technology (www.idahotech.com) have developed a portable solution for the American military that is used for onsite monitoring for the presence of biological agents, such as anthrax. The original tube Q-PCR machines were developed by ABI (www.appliedbiosystems.com) and have a range of model numbers including 5700, 7700 and the new 7900. This platform has now been licensed more widely and compatible equipment can be sourced from Biorad, Hybaid and MJ Research. These platforms will perform amplifications in approximately 2.5 to 3 hours, although they are configured to work on a 96 well plate format, lending them to high throughput sample analysis. The major restriction on this technology in addition to the limit on measurement accuracy is the fact that RNA must be extracted from tissues and converted into cDNA prior to analysis. This enzymic step adds cost and can be problematic as the enzymes involved are easily inhibited.

Hybridisation Protection Assay (HPA): This is a proprietary methodology which relies on the direct hybridisation of RNA with a chemiluminescent labelled oligonucleotide probe. The chemiluminescent molecule is linked to the centre of the oligonucleotide probe and inserts itself into the major cleft of the DNA/RNA hybrid when it is formed. Exposure of the mixture to alkaline inactivates the chemiluminescent molecules not inserted into the DNA helix, far more rapidly than those that are part of a duplex structure. This has enabled a direct homogeneous assay to be developed for specific transcripts which has the added benefit of being base pair specific (i.e. a signal is not recorded even if a single nucleotide mismatch is present). The limitations of the technology are in its sensitivity (high abundance genes and significant amount of material is required) and the fact that these tests are generated by the commercial originators of the chemiluminescent technology, only Molecular Light Technologies Ltd (www.mltresearch.com). At present, test kits only exist for the measurement of VTG from rainbow trout and fathead minnow (see Section 3.3).

Electronic gene detection: Considerable excitement has surrounded the detection of DNA using changes in conductivity. To date, a small number of commercial platforms have been developed that use this approach [eSensor (www.motorola.com/lifesciences/esensor) and Nanogen (www.nanogen.com)]. This technology provides considerable promise since the detectors are extremely cheap and relatively portable. However, their sensitivity is extremely limited and prior amplification is at present essential, therefore currently, this technology is only appropriate for detection (presence or absences) rather than quantification. Additional advantages of this technology surround its ability to detect single base changes, a property which may be suitable within specific applications.

3. HOW CAN GENOMICS BE USED IN THE CONTEXT OF THE ENVIRONMENT? – CASE STUDIES

3.1 Changes in Gene Expression in response to Eutrophic and Oligotrophic Conditions

One of the first papers to use transcriptomics in environmental microbiology investigated the impact of incubation conditions on gene expression in *E. coli* (Tao et al., 1999; the experiment design is shown in Figure 3.1). This work has subsequently been repeated within a NERC funded Industrial CASE programme by AstraZeneca Brixham, Yorkshire Water Services and the University of Newcastle.

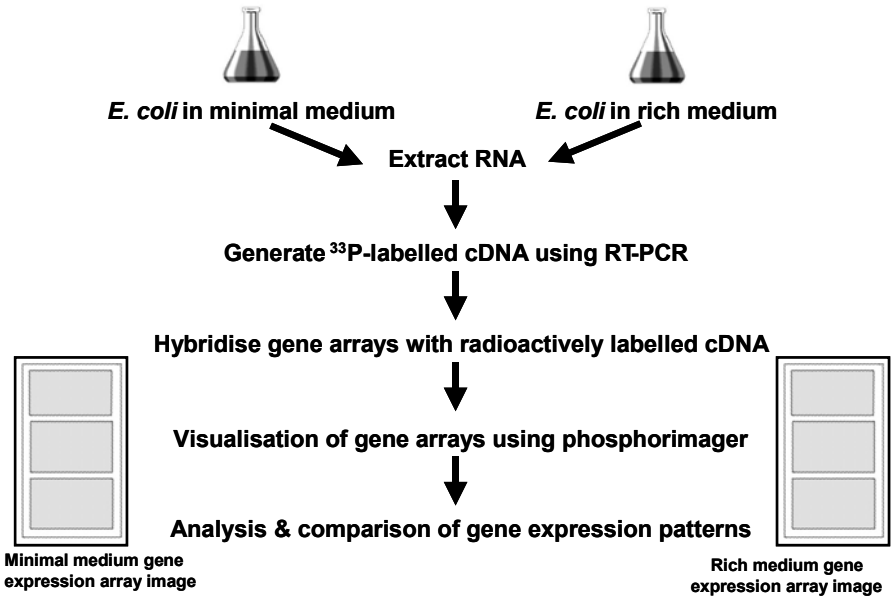

Figure 3.1 Impact of nutrient status on gene expression in E. coli

A microarray transcription profile of genes expressed in response to eutrophic (rich environment) and oligotrophic (minimal environment) incubation is shown in Figure 3.2. Genes expressed in response to eutrophic incubation appear green, those genes expressed in response to oligotrophic incubation appear red, and those genes expressed under both incubation conditions appear yellow.

These expression profiles were analysed with respect to their log ratio response to eutrophic and oligotrophic incubation; Figures 3.3 and 3.4. A log ratio of greater than zero highlights those genes that are more highly expressed in response to oligotrophic incubation, and a log ratio of less than zero highlights those genes that are more highly expressed in response to eutrophic incubation. The data in Figure 3.3 clearly demonstrates that cells of *E. coli* present in oligotrophic conditions have significantly more genes expressed at a higher level than cells incubated in eutrophic conditions. In addition, at the whole genome level (i.e. looking at the response of all 4290 genes) the results obtained from the studies by AstraZeneca Brixham and the University of Newcastle were consistent with those findings published by Tao et al. (1999; Table 3.1). However, detailed analysis of each of the 22 functional groups to which the 4290

genes belong demonstrated that whilst results between the two experiments were comparable at the whole genome level only five of the functional groups were statistically comparable; Table 3.1. Within these five functional groups only a total of 22 gene responses were statistically significant, between the two experiments, in oligotrophic conditions and only ten gene responses were statistically significant, between the two experiments, in eutrophic conditions; data not shown.

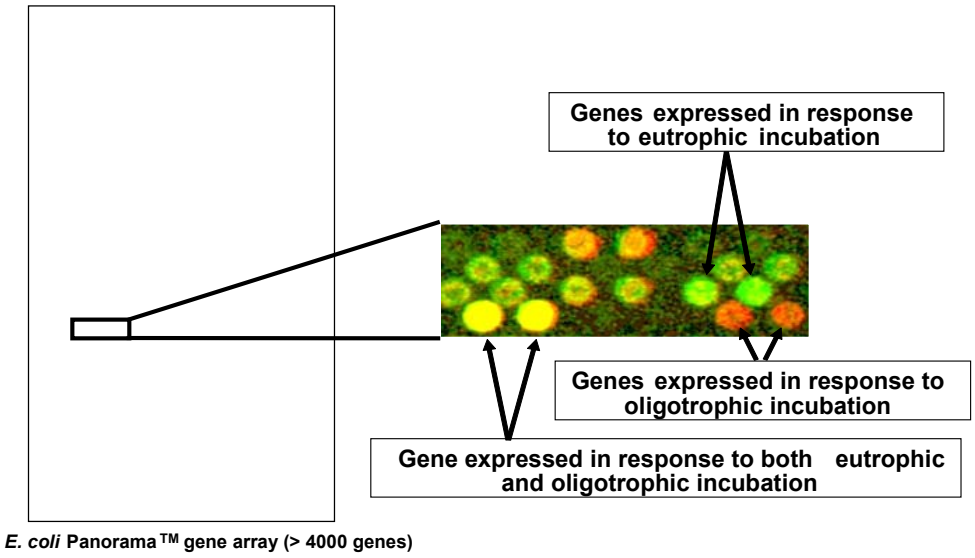

Figure 3.2 ***E. coli*** **transcriptomic response to incubation in oligotrophic and eutrophic media**

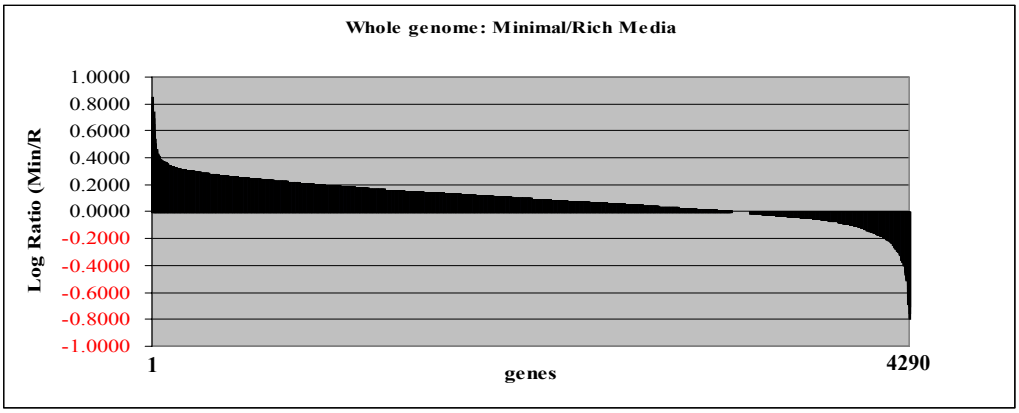

Figure 3.3 **The log normal ratio of** ***E. coli*** **gene responses for cells subjected to oligotrophic and eutrophic conditions**

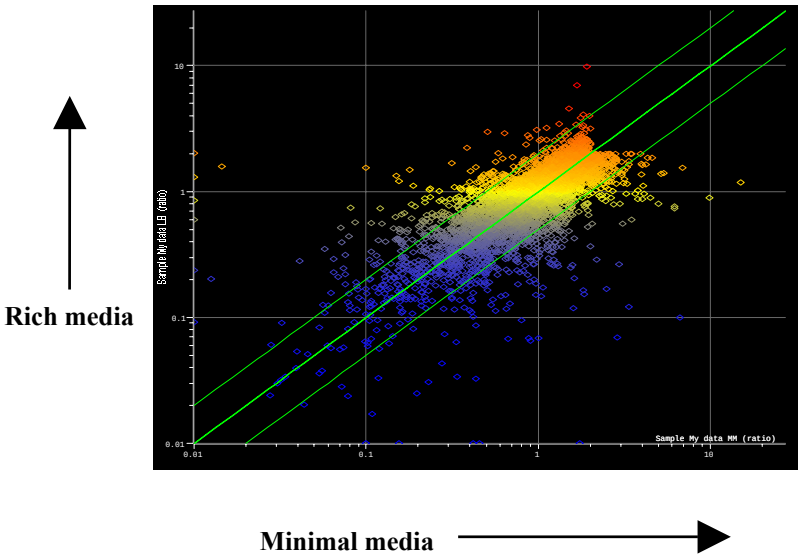

Figure 3.4 **Comparative gene response of *E. coli* cells incubated in oligotrophic and eutrophic conditions**

Table 3.1 **Statistically significant gene responses (shown in red) at the functional level for cells of *E coli* incubated in oligotrophic conditions**

Whole genome	4290	0.081	0.000*
Amino acid biosynthesis	97	0.244	0.016**
Biosynthesis of cofactors, prosthetic groups, carriers	106		
Carbon compound metabolism	124		
Cell processes	170		
Cell structure	85	0.33	0.002*
Central intermediary metabolism	149		
DNA replication, repair, restriction modification	105		
Energy metabolism	136		
Fatty acid and phospholipid metabolism	41		
Hypothetical, unclassified, unknown	1428	0.061	0.021**
Nucleotide biosynthesis and metabolism	66		
Phage, transposon, or plasmid	91		
Putative cell structure	43	0.363	0.017**
Putative enzymes	453		
Putative factors	67		
Putative membrane proteins	54		
Putative regulatory proteins	167		
Putative transport proteins	291		
regulatory function	208		
Transcription, RNA processing, and degradation	28		
Translation and post-translational modification	128	0.385	0.000*
Transport and binding proteins	254		

The above studies highlight that the same organism has a statistically different gene expression profile under extremes of nutrient status. For both studies more genes were expressed at higher levels in nutrient limited environments. This is expected as the absence of complex carbohydrates and amino acids necessitates the synthesis of cellular material from base minerals and simple sources of carbon and nitrogen. Many of the genes expressed at high levels under these nutrient limited conditions are involved in anabolism. From an ecological perspective, the fact that more genes were expressed at a higher level under nutrient stress may result in an increased sensitivity to other environmental impacts i.e. more gene targets. However, in nutrient rich conditions where the number of targets is fewer it is likely that these genes may have an increased ecological importance.

In these studies the cells of *E. coli* have clearly responded very differently to both exposure conditions, i.e. molecular adaptation in response to their environment to ensure continued growth and survival. These adaptive responses to changes in environmental conditions occur in all wildlife populations and it is important to understand the mechanistic basis of these responses. The susceptibility of an individual to other environmental impacts (pollution etc.) will differ between different exposure/ ecological contexts. The studies described above are currently being extended to determine the comparative susceptibility of *E. coli* to para-nitrophenol under oligotrophic and eutrophic conditions.

Another issue raised by the above study was the inter-laboratory comparisons. At the lowest resolution (i.e. the whole genome) the data generated by both laboratories was comparable. However, closer analyses at the functional level (Table 3.1) and the individual gene level indicated that very few functional groups and genes were statistically comparable. This may be attributed to slight differences in experimental approach that were not documented in the Tao et al. (1999) paper, or that the specific strain of *E. coli* used for both studies had in some way changed during laboratory culture. These data have important implications when considering the integration of genomics into eco-toxicology i.e. the low level of inter-laboratory comparability. Before genomics can be used within a regulatory context the issues of intra-laboratory and inter-laboratory reproducibility need to be addressed and solved. These and other published datasets indicate the importance of working to the published MIAME and MIAME/ Toxicology standards (Minimum Information About a Microarray Experiment; www.mged.org; see Section 5.2.3). Compliance with these standards may help ensure that all relevant experimental information is captured such that work can be repeated in the future without introducing variables other than the experimental scientists conducting the study and host laboratory.

3.2 Molecular biological characterisation of ammonia-oxidising bacteria within a domestic effluent treatment plant treating industrial effluents

Project Outline

The wastewater treatment plant studied as part of this research project treats domestic effluent with a significant industrial input. In order to treat this effluent a variety of different reactors configurations (activated sludge plant, trickling filter beds and a Biopur biological aerated filters) are employed and the plant is known to experience problems with nitrification. Both the industrial influent to, and the treated effluent from, this facility has an ammonia consent.

The work described below is part of a collaborative project between AstraZeneca Brixham & Newcastle University. The objectives of the study are:

To characterise and identify the ammonia-oxidising bacteria (AOB) present within the treatment plant.

To compare the microbial ecology, with respect to AOB, between the full-scale and laboratory-based reactors.

To develop a rapid molecular-based assay for quantifying specific AOB populations within wastewater treatment that can be used as a routine process monitoring tool.

To develop a nitrification inhibition test to assess the potential toxicity of process effluents to relevant species of AOB.

Provide guidance to optimise nitrification and develop recovery strategies when nitrification fails.

The remainder of this case study will focus on the characterisation of the AOB community.

Background

Nitrification is a microbiologically-mediated two-stage process. The first stage results in the oxidation of ammonia to nitrite via hydroxylamine. The second stage oxidises nitrite to nitrate. Nitrification is an important design and process consideration in wastewater treatment systems for the removal of ammonia. Ammonia has an acute toxicity to fish and other aquatic species, and can lead to eutrophication of environmental waters.

Due to the sensitivity and slow-growing nature of nitrifying bacteria, nitrification is readily inhibited by a variety of toxic components that may be present in wastewater. The mechanism and level at which inhibition is elicited is not understood and is consequently difficult to measure, predict and control. Current methods to assess nitrification inhibition are relatively time consuming and insensitive. They also rely on measurements based on foreign, unrepresentative microorganisms whose physiological properties, and susceptibility to inhibition may differ from that of the indigenous nitrifying bacteria. The development of a rapid realistic test to determine nitrification inhibition would be invaluable in the optimisation of nitrification in wastewater treatment. A useful approach to assess nitrification inhibition is to monitor the actual response of the nitrifying population present in wastewater.

Ribosomal ribonucleic acid (rRNA)-based molecular techniques provide the tools to achieve this. Molecular studies allow the direct determination of the effects of inhibiting material on indigenous bacteria. Furthermore these techniques have the potential to quantify the level of inhibition and determine whether specific inhibitors differentially affect particular members of the bacterial population. Molecular techniques such as denaturing gradient gel electrophoresis (DGGE) and differential "staining" of specific bacterial cells using fluorescence *in situ* hybridisation (FISH) can be used to characterise the ecology and population dynamics of the nitrifying bacteria.

The autotrophic AOB responsible for the initial oxidation of ammonia to nitrite are the primary focus of this work. The reason for this is that nitrification is driven by this initial step and ammonia oxidation is thought to be the most sensitive step, as nitrite rarely accumulates.

Methods

Briefly, biomass samples were obtained from the activated sludge pretreatment plant, primary and secondary trickling filter beds, and the Biopur biological aerated filter. Total nucleic acid was extracted and purified from each sample and 16S ribosomal deoxyribonucleic acid (16S rDNA) sequences amplified by polymerase chain reaction (PCR) using primers that were specific for AOB. PCR products were analysed by denaturing gradient gel electrophoresis (DGGE), cloned, sequenced and compared to 16S rRNA databases. Fluorescent oligonucleotide probes were then designed to allow the quantification of the most abundant AOB present in the biomass using FISH (Figure 3.7).

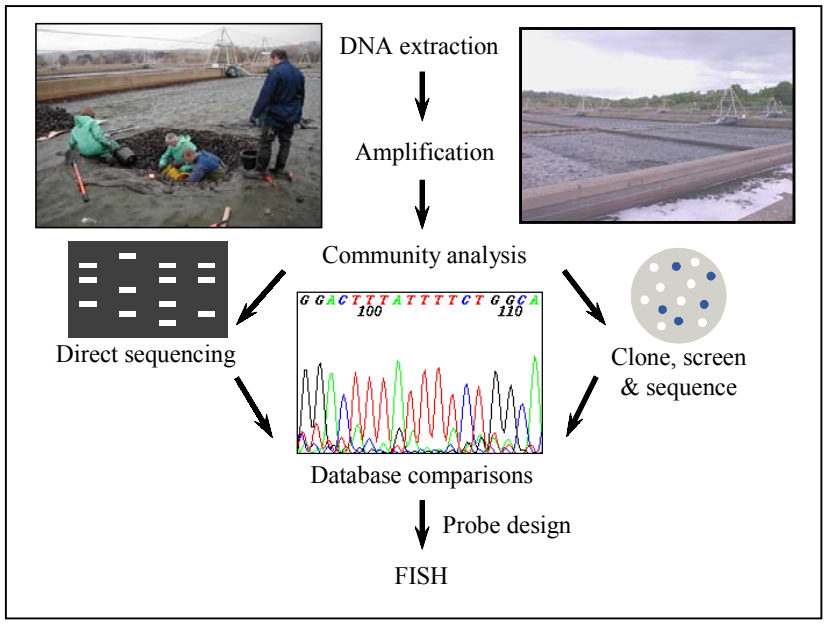

Figure 3.5 Isolation, extraction and molecular analysis of AOB populations

Results

DNA was successfully isolated from all samples (Figure 3.5). PCR using AOB specific primers and subsequent community analysis using DGGE highlighted a number of bands (Figure 3.6). Each band corresponds to the presence of at least one distinct bacterium. A wide diversity of AOB were found throughout the treatment plant with the greatest diversity in the secondary filter bed. Differences and similarities could be observed between the reactor configurations. One dominant band was observed within all reactor configurations.

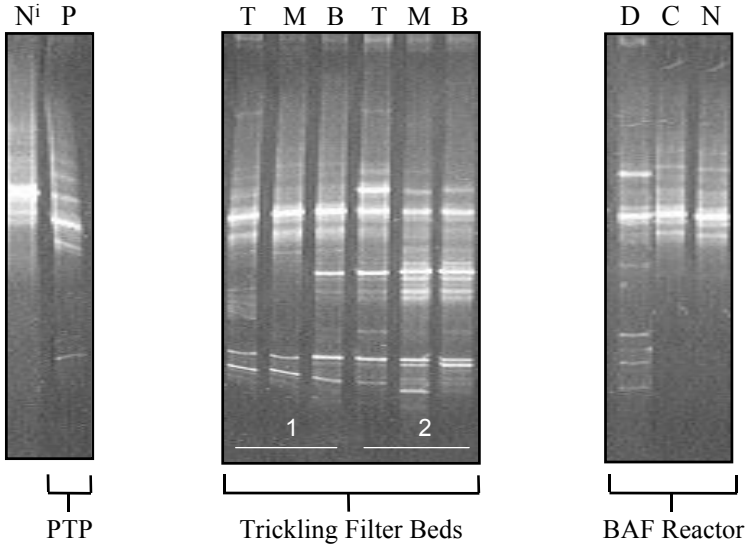

Figure 3.6 DGGE Community profile of AOB present in the treatment plant. T, M & B correspond to the top, middle and bottom of the primary and secondary filter beds; D, C & N correspond to the denitrification, carbon removal and nitrification basins of the Biopur reactor; P corresponds to the activated sludge pretreatment plant (PTP); Ni represents Nitrotox™, a commercial preparation of ammonia oxidising bacteria

The principal bacterial component of Nitrotox™, a commercial preparation of nitrifying bacteria used in inhibition studies, was identified as *Nitrosomonas eutropha*. This was not detected in any of the samples studied. This is a particular concern as toxicity-based consents are determined using this commercial bacterial preparation by the water authority accepting the industrial effluent. Cloning and sequencing of DNA sequences from the reactors confirmed that all the bacterial sequences amplified by PCR were members of the *b*-proteobacteria. The majority of the clones sequenced also clustered with the *Nitrosomonas* subgroup of AOB with the dominant AOB present in all bioreactors being closely related to *Nitrosococcus (Nitrosomonas) mobilis*. This is a halotolerant AOB that was first isolated from seawater. None of the clones obtained clustered *with Nitrosomonas europaea*, the AOB for which there are physiological data due to its ease of culture within the laboratory.

FISH studies using probes designed specifically against *Nitrosococcus mobilis* were used to detect the relative abundance of the target AOB in all bioreactor samples.

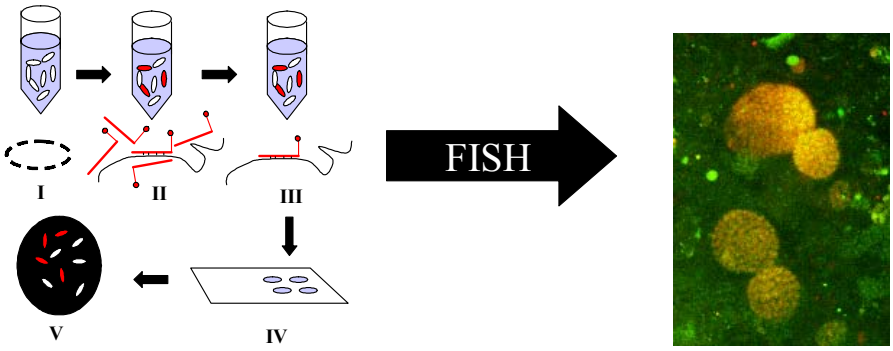

Figure 3.7 FISH analysis of *Nitrosococcus mobilis* (yellow cell clusters) in biomass taken from a trickling filter bed. I- fixation and permeabilisation; II- hybridisation; III- Washing; IV- Filtration; and V- visualization by confocal laser scanning microscopy

Conclusions

AOB are present throughout the effluent treatment plant, even in reactors where nitrification inhibition was thought to exist.

The greatest diversity of AOB populations was found in the secondary filter bed.

Nitrosococcus mobilis is the dominant AOB throughout the treatment plant. This organism is known to be halotolerant.

The AOB present in Nitrotox^{TM} is not indigenous to the treatment plant. This raises some serious questions regarding its use in toxicity assessment and bioaugmentation.

Subsequent investigations

Subsequent work has investigated changes in community structure in response to changes in the process configuration. A microtitre plate assay has been developed that can selectively capture, detect and quantify the relevant AOB characterised as part of this study. This quantitative aspect of this assay is currently being validated against FISH measurements.

3.3 Endocrine disruption in fish, with a focus on oestrogenic effects

It is now known that there are a wide range of chemicals discharged into the environment that have the capacity to disrupt the endocrine system of vertebrates and, in turn, lead to alterations in growth, development and/or reproduction (Tyler et al., 1998; Vos et al., 2000). Most of the evidence for endocrine disruption in wildlife has come from animals that live in, or are closely associated with, the aquatic environment (Tyler et al., 1998). This is perhaps not surprising

given that: (a) our freshwater and marine environments act as repositories for large volume discharges of thousands of different chemicals; (b) for wildlife associated with aquatic environments, major routes of chemical exposure include via the water/sediment - across the skin and gill surfaces-, as well as via the diet; and (c) many aquatic (and some semi-aquatic) animals deposit their eggs/embryos into the water thus exposing these vulnerable life stages directly to cocktails of endocrine disrupting chemicals [EDCs (and other chemicals)].

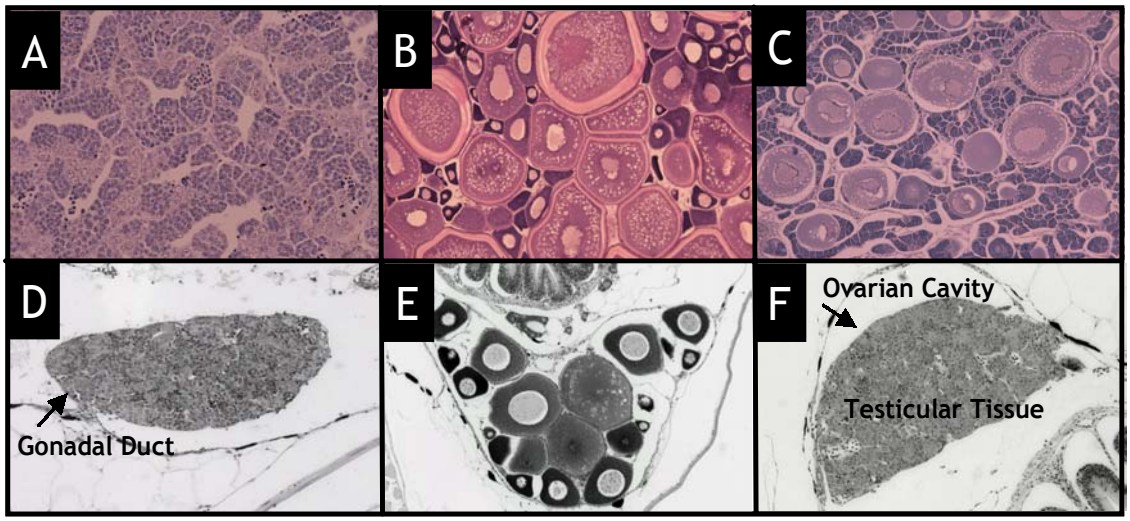

Figure 3.8 Gonadal intersex in roach (*Rutilus rutilus*). Intersex is characterised by the presence of female sex cells (oocytes) and/or the development of a female duct (ovarian cavity) in the testis of 'male' fish. Panels A-C show testis ovary and intersex gonad respectively from germ cells whilst D-F show testis, ovary and intersex gonad respectively from gonadal ducts

Much of the field evidence for endocrine disruption in aquatic vertebrates has come from studies in fish. Extensive studies on the white sucker and Lake whitefish in Canada have shown that exposure to bleached kraft mill effluents alters the endocrine status of fish, causing disruptions at the brain-pituitary-gonadal axis, inhibiting sexual development and reproduction (Munkittrick and Vanderkraak 1994). Similarly in Florida, masculinisation and behavioural changes have been shown to occur in the mosquito fish (*Gambusia affinis*) living downstream from paper and pulp mills (Bortone et al., 1989). Sexual disruption (induction of vitellogenin [VTG], altered duct and germ cell development; Figure 3.8) has been shown to be prevalent in wild populations of both roach (*Rutilus rutilus*) (Jobling et al., 1998; Figure 3.9) and gudgeon (*Gobio gobio*) (Van Aerle et al*.,* 2001) in UK Rivers and the phenomenon linked with exposure to treated sewage effluent through both field studies (Figure 3.10) and controlled laboratory exposures (VTG induction and gonadal duct disruption; (Rodgers-Gray *et al.,* 2001; Hecker et al*.,* 2002). Feminizing effects in wild fish, including intersex, have similarly been identified in Germany (Hecker et al*.,* 2002), France (Flammarion et al., 2000), Sweden, (Larsson et al., 1999) and Denmark, (Christiansen et al*.,* 2000) and again appears to be a consequence of exposure to effluents from sewage treatment works (STWs). Sexual disruption in fish has also been demonstrated in the marine environment in coastal waters in

the UK (*Platichthys flesus;* Lye et al*.,* 1997), and Japan (*Pleuronectes yokohamae*; Hashimoto et al*.,* 2000). Of perhaps greatest concern is that very recent data have shown that intersex fish (roach), that arise as a consequence of exposure to treated sewage effluents in UK rivers, are compromised in their reproductive capacity, producing sperm with poorer mobility and with a lower fertilisation success (Figure 3.11; Jobling 2002a &b) and this, in turn, potentially, has population-level consequences.

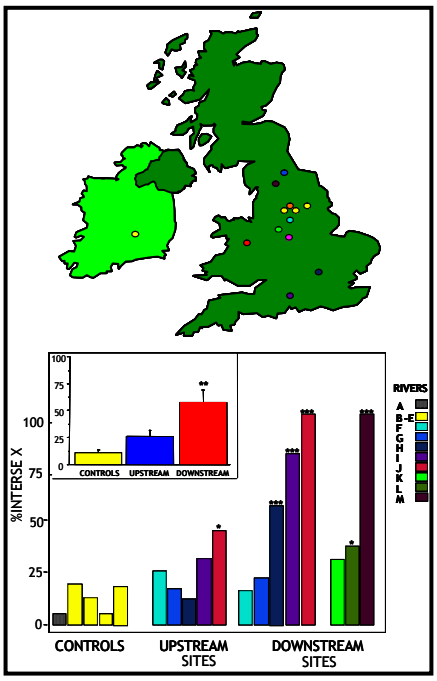

Figure 3.9 Incidence of intersexuality in 'male' roach sampled from various rivers in the British Isles. The proportion of intersex roach (containing oocytes in their testis and/or with female reproductive ducts) in rivers (F-M), lakes or canals (B-E) in England and Southern Ireland and in a laboratory control population (A). Sites B-E received no sewage treatment works effluent, whereas F-M received varying amounts of treated sewage effluent. Rivers F-J were sampled up and down-stream of major STWs (the two sites on these rivers were several kilometres apart and separated by one or more physical barriers). The inset diagram illustrates the general trends in the data when results from control, upstream and downstream sites were pooled. The asterisk denote significance from the field control sites (B-E) at the following significance levels *p<0.05, **p<0.01, ***p<0.001. Jobling et al., 1998

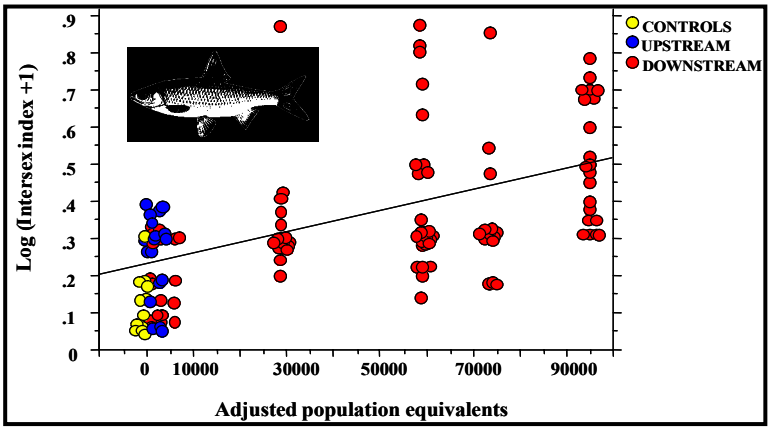

Figure 3.10 Relationship between the concentration of sewage effluent (measured as adjusted population equivalents) in river water and the degree of intersexuality in wild populations of roach throughout the British Isles (from Jobling et al., 1998)

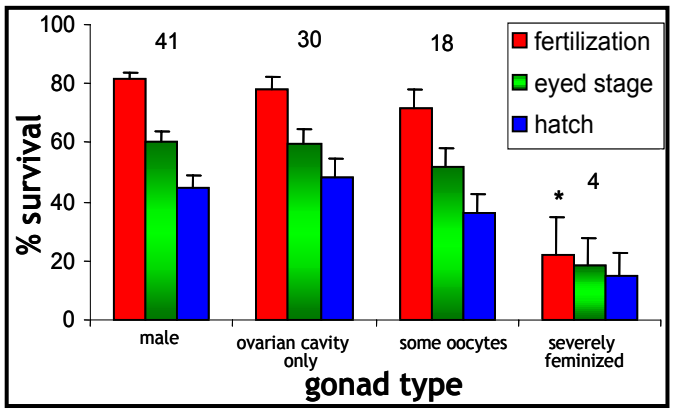

Figure 3.11 Effects of the intersex condition on reproductive success in wild roach

Most (but not all) of the feminizing effects seen in wild fish are believed, or known, to be oestrogen-mediated processes [e.g. VTG induction, the presence of an ovarian cavity in males and/or oocytes in the testis (reviewed in Jobling and Tyler, 2003)]. Some effluents in the UK (and more widely in Europe) have been shown to contain the natural steroid oestrogens, oestradiol -17β, oestrone and oestriol (all thought to be largely derived from women) and/or the synthetic oestrogen 17α-ethynylestradiol (derived from the contraceptive pill) at concentrations sufficient to cause reproductive effects seen in wild fish in UK Rivers, and for some effects (VTG induction and gonad duct disruption) laboratory studies have proven this (Desbrow et al., 1998; Tyler and Routledge, 1998; Thorpe et al., 2002). Effluents from STWs contain complex mixtures of chemicals, however, including a wide range of chemicals that mimic steroid oestrogens. Laboratory-based exposures of fish to some of these oestrogen

mimics have shown that although only weakly active, in some instances they are present at high enough concentrations in effluents to cause alterations in reproductive function (e.g. 4-nonylphenol, Thorpe et al*.,* 2002). Furthermore mixtures of steroid oestrogens and their mimics can be additive in their feminising effects (Thorpe et al*.,* 2001; Thorpe et al*.,* 2002). Add to this the fact that many oestrogens bioconcentrate (for steroid oestrogens in fish by up to 10,000-fold) which may increase any biological effect they might have, Larsson et al*.,* 1999), and it is perhaps not surprising that the primary focus on EDCs has been on oestrogens in the environment. The need to develop a better understanding of the effects of exposure to oestrogens in fish and other wildlife is illustrated by the fact that widespread screening and testing programmes have been established for assessing the biological effects of oestrogenic chemicals in Europe, Japan and the USA (Organization for Economic Cooperation and Development, Final report of the fish expert consultation meeting and US Environment Protection Agency, Endocrine Disruptor Screening Programme, Proposed Strategy Statement of Policy).

Biomarkers in fish, including molecular biomarkers have been developed to improve our understanding of the mechanisms of oestrogenic effects of EDCs and as tools for environmental monitoring for oestrogen exposure. The most notable of these in fish (and for some other oviparous animals too) is induction of VTG -an oestrogen- dependant yolk protein precursor. VTG is normally present only in the plasma of female fish, but males contain the VTG gene and can synthesise it in response to exposure to exogenous oestrogens (Chen, 1983). Useful features of VTG induction are the specificity for oestrogens, sensitivity and magnitude of the response possible [plasma VTG may increase by up to a million-fold, from ng/ml to mg/ml concentrations (Tyler et al., 1990)]. VTG induction in fish has been successfully employed to quantify the oestrogenic activity of chemicals both *in vitro* [e.g. in primary hepatocyte cultures (Gagne and Bliase, 1998)] and *in vivo* in laboratory exposures (Thorpe et al*.,* 2002a & b) and to assess the potency of oestrogenic effluents in the field (Harries et al*.,* 1999). Furthermore, field studies on wild roach in UK rivers have shown a raised content of plasma VTG in intersex fish (Figure 3.12, Jobling et al*.,* 1998). Measurement of VTG mRNA has proved to be equally effective as VTG protein for quantifying responses to steroid oestrogens and their mimics (Folmar et al*.,* 2000; Hemmer et al*.,* 2002, Hemmer et al*.,* 2001; Schmid et al*.,* 2002; Thomas-Jones et al*.,* 2003). Partial or full length VTG cDNAs have been cloned in 15 different species of fish, including both laboratory and field study species (National Center for Biotechnology Information - NCBI database). VTG mRNA can therefore now be measured in a wide range of fish, for application to both laboratory and field studies. VTG protein and mRNA, are now being employed in routine screening and testing of chemicals for oestrogenic activity. Recent studies have quantified the temporal induction dynamics of VTG and VTGmRNA for some fish species (Schmid et al*.,* 2002). VTG (including VTG mRNA) induction is presently being considered by OECD and the USEPA as an endpoint in the development of standardised tests for EDCs (see OECD and US-EPA recommendations). In the validation process for this, international ring tests are currently being undertaken to establish inter-laboratory variation for quantifying VTG/VTGmRNA. The methodologies employed for quantifying VTG mRNAs have included semi-quantitative PCR, quantitative PCR and hybridisation protection assays.

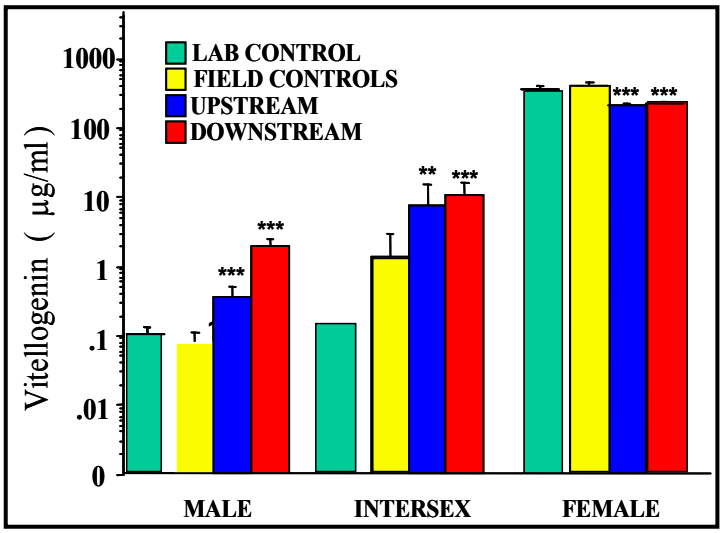

Figure 3.12 Concentrations of plasma vitellogenin in populations of roach of mixed sex in rivers, lakes and canals in England and southern Ireland and in a laboratory control population. The 'field' controls represent fish from lakes and canals that received no sewage treatment work effluent, whereas 'upstream' and 'downstream' represent data from 5 and 8 rivers, respectively, that received varying amounts of STW effluent. The asterisk denote significance (assessed by ANOVA on log-transformed data) from the field control sites (B-E) at the following significance levels *p<0.05, **p<0.01, ***p<0.001. Jobling et al., 1998

A second set of biomarkers that have been developed for detecting oestrogen exposure in fish are the vitelline envelope proteins [VEPs] (Arukwe et al*.,* 2000; Celius and Walther, 1998a & b). These glycoproteins are egg envelope components, which form the chorion of the developing egg. Three such proteins have been identified in fish (VEP1, 2 and 3). The VEPs are normally synthesised in females only [during oogenesis, starting before the onset of vitellogenesis] (Arukwe et al*.,* 2000; Celius and Walther, 1998a & b) but again males can and do synthesise VEPs in response to exposure to exogenous steroid oestrogens and their mimics (Arukwe et al*.,* 1997). The liver has been shown to be the source of these VEPs in the rainbow trout, medaka (*Oryzias latipes,* Hamazaki et al*.,* 1989), sea bass (*Gadus morhua*) and Atlantic salmon (*Salmo salar*, Oppen-Bernsten et al*.,* 1992). In some teleosts, the oocyte/follicle is involved with the synthesis of the zona pellucida (e.g. pipefish, *Sygnathus fuscus,* Begovac and Wallace 1989), carp (*Cyprinus carpio*, Cotteli, 1998) and goldfish (*Carassius auratas,* Cotteli, 1998), as occurs in amphibians and other mammals. VEPs are extremely hydrophobic and difficult to measure the protein using conventional techniques. In contrast, the mRNA for these targets can be measured with ease for the cloned sequences. VEP mRNAs are equally responsive to oestrogen stimulation as VTG and possibly more so (Celius and Walther, 1998a). Partial or full length VEP cDNAs have been cloned from 8 species of fish (NCBI database). The same methods used to quantify VTG mRNA have been used to measure VEP mRNA transcripts (semi-quantitative PCR, quantitative PCR and hybridisation protection assays).

Other genes central in the oestrogen response pathway that have been studied with a view to unravelling the mechanisms of oestrogenic disruption in fish include oestrogen receptors (α, ß and γ), various enzymes involved with sex hormone biosynthesis (including, aromatase and C17 lyase), and gonadotrophins (follicle stimulating hormone and luteinising hormone; hormones that are derived from the pituitary gland and control gonad development and sex steroid synthesis). *In vitro* studies using fish hepatocyte cultures have shown that steroid oestrogens and their mimics affect oestrogen receptor mRNA regulation (Flouriot et al*.,* 1995). *In vivo* studies have shown that exposure to exogenous steroid oestrogens and the aromatase inhibitor fadrozol alter gene expression of both brain and gonad aromatases (a key enzyme in the biosynthesis of oestrogens from androgens). Exposure to pharmacological doses of fadrozol during the window of sexual differentiation in the zebrafish results in a depressed aromatase mRNA expression and subsequently in all-male fish populations [Fenske – personal communication]. Exposure to steroid oestrogens in adult fish similarly alters gene expression of brain aromatase, but the significance of this for reproduction has not been determined (Halm et al*.,* 2002). Recent studies have shown that exposure to the oestrogen mimic nonylphenol disrupts pituitary synthesis of gonadotrophin mRNA, and thus disrupts gonadal development, at a pivotal level in the endocrine control pathway for sexual development. These examples serve to illustrate that molecular biology, specifically the study of single specific gene transcripts, is playing an increasingly important role in unravelling the pathways and mechanisms of oestrogenic disruption in fish.

Endocrine disruption in fish (and for other wildlife) is a complex issue. Most of the focus on endocrine disruption in fish has been on chemicals that are and/or mimic oestrogens, largely because of the original finding of feminized fish in UK rivers. However, there are examples of other types of endocrine disruption in wild fish populations, for example, androgenisation of mosquito fish living downstream of paper mill factories in Florida, USA (Munkittrick and Vanderkraak, 1994). Indeed, it is now well established that chemicals that mimic hormones are not restricted to 'oestrogens', but also include mimics of androgens, progestagens, thyroid hormones, corticosteroids and their antagonists (Tyler et al., 1998). Furthermore, some chemicals can act at multiple targets to disrupt physiological function (e.g. nonylphenol acts as an oestrogen agonist, androgen antagonist and can alter gonadotrophin synthesis and secretion, Scholz and Gutzeit, 2000). Given that EDCs (and other chemicals) can have combination (e.g. additive) effects and can also interact with the nervous system to alter development, reproduction and growth, the development of more comprehensive molecular approaches (beyond assessing effects at single gene targets) is needed if we are to more fully appreciate the interactions of chemicals with the endocrine system and identify pathways and mechanisms of endocrine disruption. Projects are ongoing in the UK and more widely in Europe, Japan and the USA, to develop fish macro arrays and microarrays with this challenge in mind [see *UK Programmes*, Section 4.1].

3.4 Exploiting a Genomics Approach to Develop a Terrestrial Biomarker for Heavy Metal Contamination

Understanding the risks associated with environmental contamination with specific heavy metals such as copper, cadmium, lead and mercury has been a subject of investigation throughout the last century. All metals, whether biologically essential or otherwise, pose two fundamental problems: they are more-or-less toxic above certain threshold concentrations and they are not biodegradable. In the field of human health, metals continue to cause major

diseases, from copper causing Indian Childhood liver cirrhosis (Sethi et al., 1993) to exposure to cadmium in drinking water or food resulting in Itai-itai disease (Nogawa., et al. 1996). In addition to clinical considerations/implications, transfer and bioamplification of metal ions within food chains has been shown to have significant impacts on ecosystems. Acute incidences such as the Aznalcollar mine disaster in Spain and the Wheel Jayne mine outflow in the UK (Meharg et al., 1999; Somerfield et al., 1994), coupled with realisation of the scale of chronic long-term environmental changes around refineries such as at Kola Peninsula, NW Russia (Rigina et al., 1999) Palmerton zinc smelter, USA (Li et al., 2000) and Avonmouth cadmium/lead/zinc smelter, UK (Hopkin et al., 1989), indicate the environmental harm that can result after the release of potentially toxic metals. Therefore, when evaluating environmental emissions, assessing risk from contaminated land or legacy sites, and determining chronic exposures due to changes in waste management practices, both through landfill disposal and sewage sludge processing (Morgan, 2003), the heavy metal dimension remains a primary consideration.

However, assessing the risk posed by heavy metals is complicated by a suite of mitigating factors that influence metal bioavailability and therefore the relative toxicity. It is now well established that the obscured toxicity of a metal ion is not solely controlled by the local geochemistry (e.g. pH, clay content, organic matter in soil; hardness and competing ion in freshwater salinity in estuaries and seawater), but is also related to the ability of the organism to accumulate the metal and the cellular processes that are present to chelate the metal ions within a cell. Plants and animals are not passive metal receptors; they actively respond to environmental stresses they are exposed to in their life-cycle.

Land surveys have, to date, mainly used subjective criteria, such as previous use or geochemistry, to assign risk to heavy metal contaminated sites. In fact, it was stated within the 1988 Contaminated Land Survey of Wales that objective assessment of contaminated sites was impractical (Welsh Office 1988). Therefore, the question we posed was: could a genomics approach supply an objective marker for metal contaminated land supplying not only a measure of environmental risk but also linking individual responses to the population-level impacts? In other words, could functional parameters of free-living organisms provide quantitative measures of the cocktail of poorly-understood interactions leading to metal-toxicity.

Preliminary work has focused on identifying a marker for cadmium toxicity using a resistant population of the established terrestrial sentinel earthworm, *L. rubellus*, inhabiting a Roman mine site. Initially, classical protein biochemistry was used to identify a cellular constituent responsible for binding the bulk of the cadmium constituting approximately 1% of the body burden in this population. When purified, this protein displays many features, spectral and amino acid composition, associated with a family of well characterised, small, cysteine rich, metal binding proteins, the metallothioneins (MTs, Stürzenbaum et al., 1998). However, this approach failed to reveal further details of this potential marker. A molecular approach exploited specific metal binding motifs associated with chelation of cadmium. This allowed a genetic fingerprint to be generated of genes containing this motif from control populations and compared with that exhibited by worms exposed to cadmium (Figure 3.13). Isolation and characterisation of the major genetic element of this fingerprint, which was specific to the cadmium-exposed organisms, revealed it to encode earthworm MTs. Using this fragment from the genetic fingerprint, two MT isoforms were isolated from earthworms (Figure 3.14). Subsequently it was confirmed that the gene encoded by MT2 represented the protein isolated

from the earthworms within the initial phase of the study (Morgan et al., 2003). Furthermore, by expressing both isoforms within bacterial cells, substantial material could be generated which facilitated the production of a polyclonal antisera, that had been used to demonstrate the involvement of MT2 in the uptake, sequestration and excretion of cadmium within the earthworm (Figure 3.15), whilst MT1 displays no direct involvement in cadmium detoxification (Stürzenbaum et al., 2001). Having demonstrated that MT2 is mechanistically linked to cadmium sequestration, it was important to validate it as a quantitative marker of Cd exposure and to link it to lifecycle alterations caused by metal toxicity. The rational behind the latter approach is that genetic changes must precede and, therefore, predict whole-organism and population level (i.e. ecologically relevant) functional levels.

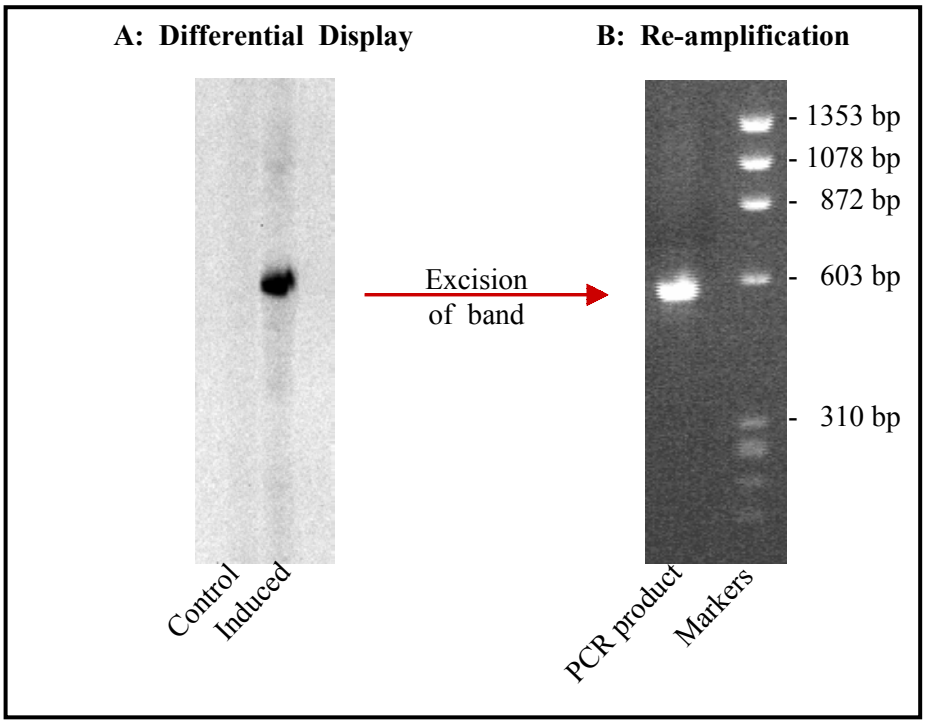

Figure 3.13 Differential display and re-amplification of earthworm metallothionein. Panel A shows the band that was excised from the differential display gel of mRNA derived from unpolluted control (Dinas Powys) and Pb/Zn/Cd induced (native Rudry) earthworm populations using a specific Cys-Lys-Cys primer and an anchored poly-T-XG primer (see[19,20]). The product was re-amplified utilizing the same poly-T primer and specific primer. The band shown is the cloned product, subjected to electrophoresis on a 2% agarose gel, containing ethidium bromide (0.5 μg/ml). Nucleic acids were visualized under ultraviolet light (panel B). Marker=nucleic acid molecular weight markers (*Hae*III digested ϕX 174 DNA, Promega)

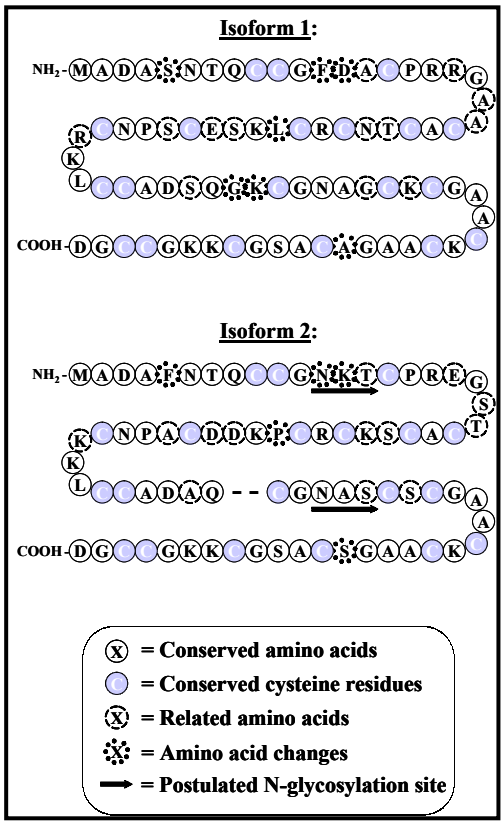

Figure 3.14 Schematic diagrams depicting the amino acid sequence of earthworm metallothionein isoforms 1 and 3.4

Preliminary validation involved laboratory exposures of earthworms to cadmium using the recommended OECD exposure protocol, but harvesting organism exposure to provide a complete temporal profile of MT2 expression. These studies confirmed that the expression of MT2 showed both time and dose-dependent responses to cadmium (Figure 3.16). Subsequently, deployment of this marker within a range of tests including laboratory exposure to "natural soils", mesocosm exposure and *in situ* exposure at abandoned industrial sites, have demonstrated the specificity and dynamic range of this marker. Furthermore, correlations of MT2 expression with parameters at higher levels of biological organisation can be determined by evaluating body burden and lifecycle parameters from the same organisms. Most striking is the correlation between MT2 expression and cocoon number, a factor used as a major sub-lethal effect parameter (Figure 3.17). In order to further evaluate the association between MT2 expression and the population consequence of cadmium exposure a demographic model for the earthworm was constructed and measured in parallel with MT2 gene analysis. This demonstrated a significant correlation between MT2 expression and the intrinsic rate of population increase. However, although a mechanistic link has been unequivocally demonstrated between body cadmium levels and MT2 expression, as yet no direct mechanistic evidence links MT expression with reproduction. However, a precedence for this link exists in fish, where MT interacts with the systems responsible for the generation of the egg yolk protein vitellogenin (Valencia et al. 1998).

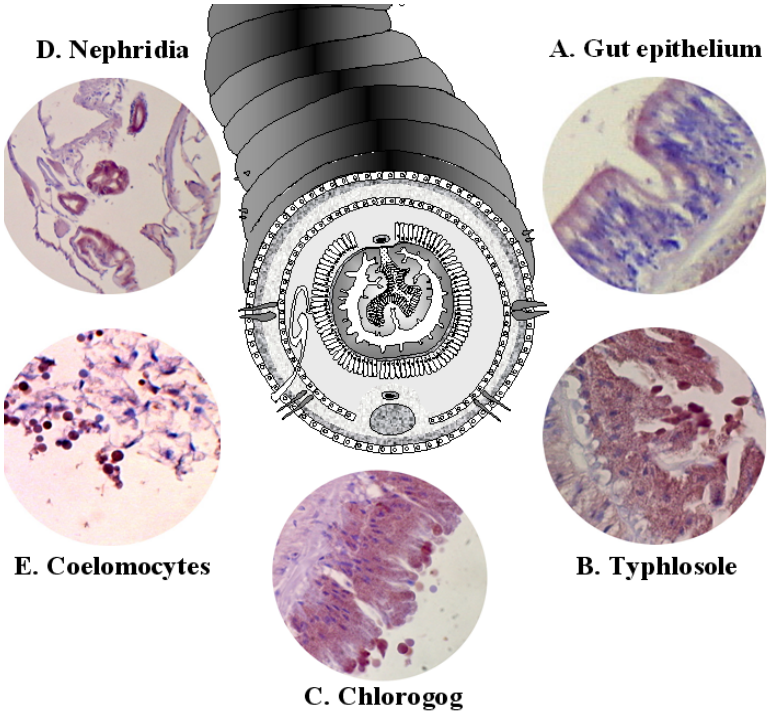

Figure 3.15 Immunoperoxidase-histochemistry performed on posterior transverse sections of the earthworm revealed the precise cellular localisation of MT2, the Cd-responsive MT. Staining was observed in the apical cytoplasm of the intestinal epithelial cells (Panel A) and in the chloragogenous tissue (panel C) surrounding the basal layer and the typhlosole (Panel B). Higher magnification of the chloragogenous tissue indicated that the MT-positive staining was most intense either within discrete vesicles in the apical cytoplasm of intact cells abutting the coelomic cavity, or as apical projections released into the coelomic cavity or typhlosole (Panels B and C). The apical projections released into the coelomic cavity were observed being engulfed by coelomocytes (Panel E) and heavy staining was observed surrounding the Nephridia (Panel D)

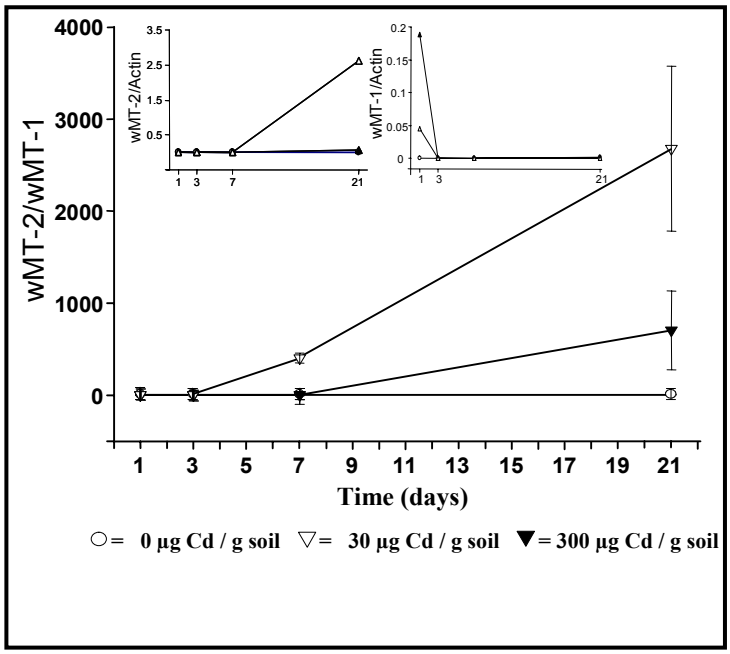

Figure 3.16 Amino acid alignment of the two major MT isoforms in the earthworm *L. rubellus*. A, between the two isoforms, all cysteine residues are conserved, and the primary sequence displays an identity of 74.7% and a similarity of 91.1%. *B*, isoform-specific transcript analysis of wMT-1 (*i*) and wMT-2 (*ii*). Quantifications were performed on an ABI PRISM® 7700 and normalized with actin. Triplicate measurements were taken of four individual earthworms sampled from a temporal- and dose-exposure experiment. The *main graphical output* in *panel B* elicits the relationship between MT1 and MT2 when exposed to cadmium over a time and dose course/regime. MT2 is the sole cadmium-responsive isoform

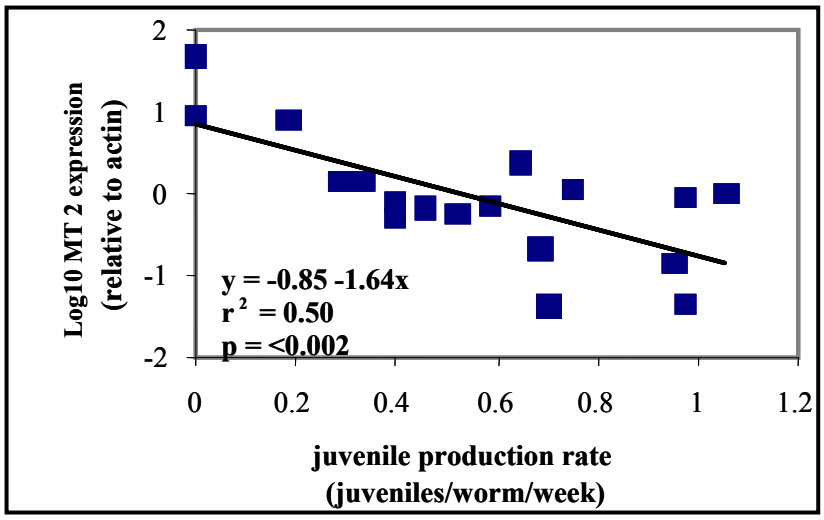

Figure 3.17 Correlation of the expression of MT2 with the juvenile production rate during exposure to cadmium

3.5 Toxicogenomic Profiling of Bioreactive Particles within Diesel Exhaust

Epidemiological studies conducted first in the USA and later in the UK, suggested that a relationship exists between increasing cardio respiratory hospital admission, morbidity and mortality rates and increase in the small particulate matter, PM10s, originating from fumes such as diesel exhausts (Anderson et al., 1991; Schwartz, 1994). PM10s are particulates with an aerodynamic diameter of less than 10 μm, they are complex mixtures of natural materials, metals, carbonaceous components, soluble ionic species and organic micro-pollutants (Richards, 1997). In urban environments, diesel exhaust particles (DEP) form a large constituent (20-80%) of the airborne PM10s arising from vehicular activities in the UK (DoH, 1995). The exact biological mechanism by which these molecules elicit their toxicological effect is unclear. Furthermore, the impact these particles exert on the wider environment is poorly understood. Previous research has concentrated on the size and surface chemistry of the PM10s, as an important factor underlying the health problems associated with exposure to these particles (Osier and Oberdörster, 1997). These studies have provided a limited insight into the biological active constituents. Therefore, it has been necessary to perform classical toxicological exposures to study the physiological or the cellular damage caused by PM10s on the lung surface (Murphy et al., 1998). Although this provided classical histopathology associated with lung damage (Richards et al., 1999), the chain linking exposure to effect was still not apparent. These results suggested that a better understanding of how DEP may increase lung permeability/information could come from studying more subtle biological endpoints. To this end, a toxicogenomic approach was employed to profile the genes involved in the toxicological response (Reynolds and Richards, 2001).

A multidisciplinary group from geologists to toxicogenomics was formed to study ambient particulate matter. To establish the linkage between particle chemistry, genomic response and toxicological endpoint all parameters were measured within each experiment. To ensure the samples were of environmental relevance, a massive sampler (supersucker) was used to collect particulates at 6 sites in the UK, these were then used for:

(a) chemical analysis

(b) cellular studies (mostly to look at oxidative metabolism)

(c) *in vitro* toxicology studies with lung epithelial target cells (human and animal primary cells)

(d) *in vivo* lung toxicology following instillation (direct deposition via the trachea under anaesthesia).

The rational was to look for genetic markers or candidate genes to monitor the toxic response. For this a commercial macroarray, representing genes involved in stress, cell signalling, xenobiotic metabolism, DNA repair/cell cycle, inflammation etc was employed. This stress macroarray includes 207 rat cDNAs, double-stopped on a positively charged nylon membrane. In negative controls, housekeeping genes are included to ensure hybridisation specificity and as positive controls. The procedure is summarised in Figure 3.18. Identical membranes were hybridised to radiolabel targets generated from the genes transcribed within control and PM10 exposed animals. After high stringency washing, the quantity of hybridisation was analysed using phosphorimaging. This allowed the relative level of each transcript to be determined during a time course of exposure to different environmental PM10

samples. Of particular interest was the water soluble component of PM and specifically its bioavailable transition metals. It was hypothesised that these are the problem agents and by being bioavailable are crossing the lung barrier and at first pass hitting the heart. The macroarray technique was instrumental in identifying 15 genes of current interest (between 8 and 5000 fold change) that alter in the heart 4 hours after the pollutant is given to the lung. This might explain why people exposed to air pollution die of cardiac problems. The studies are being extended to a range of body tissues and are expected, within a few years, to identify the most important transition metal in the mixture (believed to be zinc). More importantly, this technique is revealing how the heart talks to the lung and how relevant drugs may compromise lung defences and modulate the affects on the heart. The global implications of this work are reflected in collaborative ventures with Hawaii, Moscow, Beijing, Imperial College London, Kings College, London, Sydney and Christchurch.

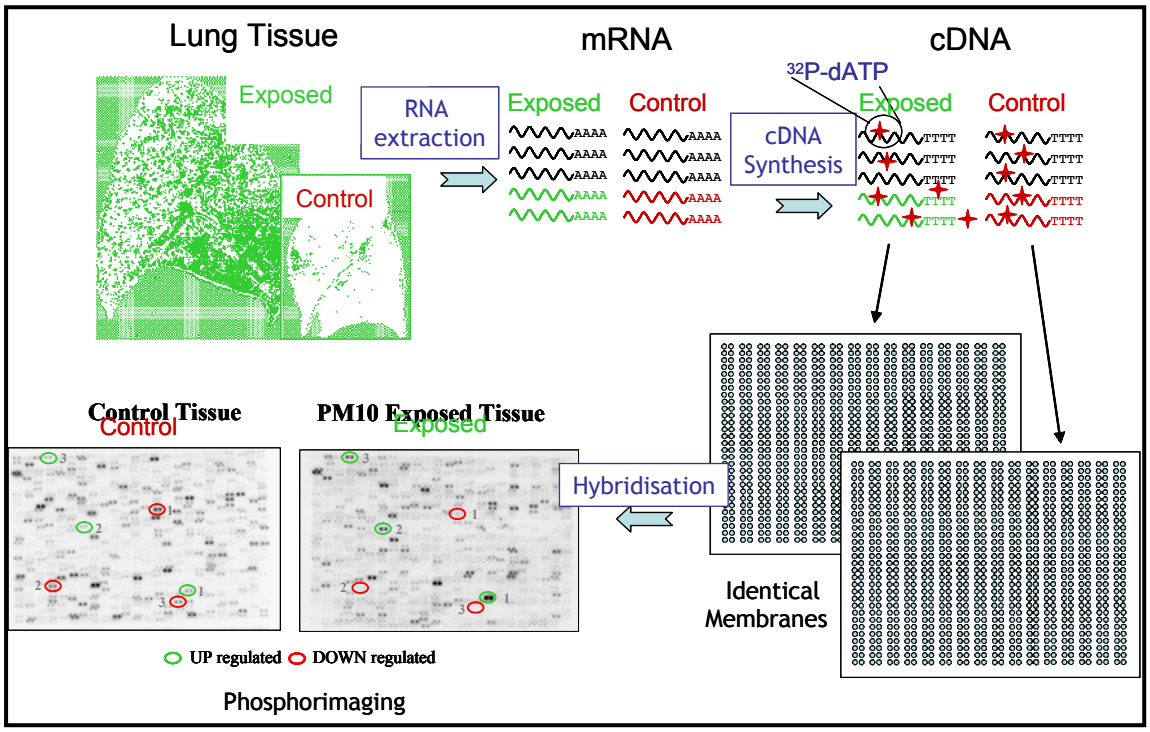

Figure 3.18 The procedure for analysis of the expression profile of multiple genes in lung tissue using a macroarray

New projects have now moved into proteomics (for *in vivo*) looking for biomarkers in lung lavage fluid related to specific biological endpoints of toxicological damage (inflammation, lipoproteinosis). Although the implications to environmentally relevant organisms can be extrapolated since the pathways involved are common, direct confirmation of the long term consequences of PM10 release into the environment on specific ecosystems requires further investigation. The tools provided by this toxicogenomic approach should enable these experiments to be performed in short order.

4. A REVIEW OF LARGE ENVIRONMENTAL GENOMICS PROGRAMMES

4.1 UK Programmes

4.1.1 NERC Environmental Genomics Programme

The NERC Environmental Genomics programme will apply genomics to the natural environment, using sequence data to advance and test evolutionary and ecological theory, and so provide a better understanding of ecosystem function in the context of biodiversity.

The NERC programme aims to advance the understanding of how organisms perceive change (at the molecular and genetic level) and make functional responses within their local environment, and to what extent variation in these responses is adaptive. These measures may be short-term adaptations, quorum sensing, regulated gene expression, survival, reproduction etc. Alternatively, they may be longer-term adaptations in respect to horizontal gene flow, community diversity, species abundance and the evolution of speciation. Genomics also allows investigation of adaptation at the level of individual versus community diversity.

The NERC Environmental Genomics programme aims to facilitate the exploitation and, where appropriate, acquisition of whole or partial genome sequence data in environmental biology. Central to this aim is the establishment of a community of researchers, comprising empiricists and theoreticians, who are proficient in the development and application of post-genome technologies to the study of organism and ecosystem function. The integration of molecular genome science with ecology and evolution will generate exciting and novel interdisciplinary research opportunities. The strategic objectives of the NERC programme and the projects funded in rounds one and two are described below. Clearly a number of these strategic objectives and proposals are immediately relevant to the Environment Agency. However, a number of projects that do not immediately appear relevant will provide valuable baseline data against which environmental impacts can be determined.

STRATEGIC SCIENTIFIC OBJECTIVES

Phenotypic and genotypic bases of fitness

Whole genome sequence, in conjunction with technologies for functional analysis, enables ecological performance (fitness) to be studied using "bottom-up" (genes to population) strategies. Objectives include:

Identification of gene loci affecting ecological performance.

Functional characterisation of traits affecting ecological performance.

Individual, population, community and ecosystem responses to the environment.

The extent and significance of allelic variation among traits affecting ecological performance.

Ecological significance of molecular variation

Chip and other technologies enable variation among populations to be quantified rapidly at multiple loci. Objectives include:

Determining the consequence of variation in the levels of gene expression at multiple loci on ecological performance.

Defining the extent of spatial and temporal variation at regulatory versus structural loci.

Identifying numbers of loci (QTLs) and their genomic distribution.

Determining community structure in unculturable organisms and investigating the response to environmental change.

Integration of genotype and phenotype

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. Objectives include:

- Responses of genomes to genetic and environmental stimuli.
- The impact the environment plays on the phenotype.

Phenotypic evolution

The relationship between DNA sequence, form and function is poorly understood. Whole or partial genome sequence, in combination with the tools of functional genomics, offers opportunities to develop a better understanding of phenotypic evolution. Objectives include:

- Explaining (at mechanistic levels) the evolution of novel traits.
- Determining rules of adaptive evolution at a theoretical level.
- Understanding the contribution of regulatory pathways to adaptive evolution.

Genome architecture

The position of genes within genomes, their number, arrangement and the nature of their interactions determine the expression of phenotypic traits. Objectives include:

- The stability of genetic architecture.
- Organisational constraints.
- Responses of genome architecture to selection.

4.1.2 NERC Post-genomic and proteomics programme

As a result of the 2002 spending review NERC have been allocated an additional £12M to invest in a post-genomic and proteomics programme. This will build on, but will be separate from, the existing NERC Environmental Genomics programme (see above). NERC is currently asking the UK environmental sciences community for suggestions for strategically important areas of research that could be addressed under this new programme. It is important to identify why the science area proposed is of strategic importance to NERC and the UK now, and in the future.

Through the £16.5M currently invested in the Environmental Genomics programme, NERC has maintained a broad scientific remit to enable post-genomic science to be integrated across many areas of NERC science. With the availability of an additional £12M from the 2002 Government Spending Review, NERC will use the present capacity-building as a springboard to focus on specific scientific areas in which the NERC community can take a leading role on the international stage. NERC has therefore opened a consultation with its research community to identify the specific areas where NERC can derive maximum impact.

NERC has requested suggestions for such strategically important areas of research that could be addressed using an integrated genomics approach. In doing so NERC have highlighted some example areas for potential investment (see below). NERC has only offered these as suggestions and will give equal consideration to cases received either in these or other areas.

- Establishing a coordinated UK community resource to facilitate the integration of post-genomic science, at the transcriptomic and proteomic level, into fish biology.

- Providing a bridge between biodiversity and ecosystem function, especially with respect to the biogeochemical cycles.

- Supporting the development of techniques and technologies to maximise the utility of proteomics in an integrated genomic approach to study the response of wildlife populations to environmental change and disease.

- Support for post-doctoral research fellowships to champion environmental genomics within specific areas of the NERC science strategy.

- Continued support for the Environmental Genomics Thematic Data Centre and bioinformatics.

The consultation will end on March 7th 2003 and an advisory panel will assess the consortia suggestions and, from these, they will identify three or four main areas to target under the new programme. An open call for applications against these areas will then be issued in May 2003. These applications will be assessed by a Steering Committee appointed to the new programme.

4.1.3 BBSRC - Funded Post Genomic Research

BBSRC are funding genomic-based science to revolutionise our understanding of the contribution of genetic factors to the characteristics and properties of all living organisms,

from the molecular level to cells, whole organisms, and even populations. Genomics will also advance our understanding of the evolution and interrelatedness of species. A major finding of modern biology is that genes, and their location along chromosomes, are remarkably similar in distantly related species. Because of this genomics studies in a few "model", or representative, species can provide results that can be applied rapidly to other species - for example results from research into fruit flies and the nematode *C. elegans* can be applied to many other invertebrate wildlife populations and studies on zebra fish can be applied to other fish populations. It is this comparative genomic approach that will have importance for the Environment Agency.

Genomics will make a major impact on all biology-based industries including healthcare and pharmaceuticals; animal health; crop and livestock breeding; and the use of microorganisms and plants to produce high value raw materials. However, genomics also has the potential to provide improved certainty and relevance associated with the environmental impact assessment.

In 1999-2000 BBSRC allocated £30M for genomics research. This support is in the form of major research initiatives specifically in genomics, grant awards for the application of genomics to strategic research programmes, and individual grants for developing genomics resources (functional genomics technologies, investigating gene function (IGF), gene technologies underpinning healthcare, and genomics in animal function). BBSRC is establishing specialised training in techniques underpinning genomics, both for academic scientists and bioscientists working in industry. The spending review in 2000 allocated a further £25M to exploiting genomics to deliver strategic science from the infrastructure provided through the IGF programme. A further £49M has been allocated to BBSRC in the 2002 spending review to be allocated to post-genomic science and proteomics. Research projects funded by BBSRC under all programmes other than the 2002 allocation are summarised below.

The IGF initiative was launched in 1999 to fund consortia to develop access to high throughput genomics technologies (e.g. micro-arrays and proteomics) and associated resources (e.g. mutant populations) for communities working on organisms key to the BBSRC remit.

The opportunities which the availability of sequenced genomes offer for the rapid identification of gene function are transforming biological research and its applications. As such BBSRC widened the access to these developing technologies that underpin this revolution to stimulate research which will fuel the growth of the UK knowledge and skill base in this field, promote broadening awareness of the potential of the techniques and stimulate their application to major problems in key organisms.

The IGF initiative was the second phase in BBSRC's programme to support genomics research, following the identification of increased investment in genomics as a priority in the Comprehensive Spending Review Settlement. The first phase was the Technologies for Functional Genomics initiative that sought to promote the development of tools and technologies to enable the systematic study of gene function.

The objectives of the IGF initiative are:

- to make available methods and resources whereby the connection between genes and important functions can be discovered using genomics, through providing access to e.g. microarrays, filter arrays and proteomics;

- to support the creation of new mutant libraries and similar essential resources, specialised screening of new and existing libraries and distribution of mutant collections;

- to increase focus and coordination in the community by supporting collaborative and systematic approaches to function search and gene identification within consortia based on model organisms or key organisms of commercial significance which are central to the BBSRC mission.

The outputs of the IGF initiative included:

- Increased knowledge of genes and their functions in model organisms and key organisms of commercial significance;

- Increased access to high throughput technologies (e.g. microarrays and scanners) and associated proteomics, based around e.g. central provision within a research institution (or via industrial contract in the case of high-density work), a distribution network and separate "readers" at distributed sites;

- Increased focus and coordination within the community, based on systematic, consortium-based approaches to function search and gene identification targeted on model and key organisms;

- An enhanced archive of materials and data accessible to the broadest UK possible community;

- Enhanced output of trained postdocs and technicians underpinning functional genomics research;

A small number of awards were made, focused on major model organisms and species representative of key groups of organisms of economic significance. These awards were made on the basis of a presented case that demonstrated:

- the significance of the organism either as a model widely accepted within the community as a focus for study and giving insight into major areas of science within the BBSRC remit, or as an organism (or representative of a group of organisms) of key UK economic significance to one or more of the user groups underpinned by the BBSRC Mission;

- a broad, community-based strategy for the proposal, based on wide access and complimentary and coordinated activity, to give maximum availability and derive maximum research value from facilities provided and resources being generated;

- appropriate proposals for the operation of a service, based upon consultation of and ongoing involvement by the community concerned.

As a result BBSRC concentrated its funding based around the following key representative and industrially important species:

- *Arabidopsis* (Thale cress)
- Brassicas (cabbage family)
- cereals
- *Drosophila melanogaster* (fruit fly)
- Streptomyces (antibiotic producing bacteria)
- farm and veterinary animals
- yeast
- *Physcomitrella patens* (moss).

The principal technologies and resources supported by these BBSRC IGF-funded centres include: libraries of genes and varieties; protein identification and analysis; extraction, identification and separation of metabolic products in cells; "knock-out" approaches to correlating the loss of a gene with changes in an organism's character or behaviour; and, most significantly, DNA chips or microarrays that provide a "snapshot" of gene activity in a cell at a particular time.

Genomics research funded by BBSRC as part of all their post-genomic initiatives are summarised below. Access to a full abstract describing projects that are relevant to the Environment Agency can be obtained by clicking on the title of the proposal in the electronic version of this document.

4.1.4 MRC Funded Post-genomic Research

It is both clear and obvious that MRC-funded post genomic research will have a medical bias and will focus on exploiting knowledge contained within the human genome sequence and a few other mammalian model species (e.g. mouse). Whilst the majority of their funding is targeted towards the identification of new therapeutic targets, and understanding the mechanisms that underpin disease some of their research efforts will provide techniques and approaches that can be used to assess non-occupational impacts on man and wildlife populations. Perhaps the two main areas of direct interest to the Environment Agency are the MRCs efforts in population based studies, examining multiple risk factors including environmental risk factors, and the initiatives associated with communicating the impact of these new technologies to the public. The themes and research priorities being addressed by MRC are summarised below.

Genetic information and biological function

Analysing the details of the biological processes that underpin health is a key step in the search for new ways of treating disease. Universities will have important opportunities to acquire leading edge facilities for large-scale screens, sequence and expression analysis, sample banking, and bioinformatics through the Science Research Infrastructure Fund. MRC's main efforts will be devoted to long-term investments to build up the UK infrastructure supporting leading research in these areas. Many of these investments in infrastructure are being co-supported by BBSRC. The following topics are MRC's priority areas:

Theme 1: Genetic Information and Biological Function

Analysing the details of the biological processes that underpin health is a key step in the search for new ways of treating disease. MRC's main effort is devoted to long-term investments to build up the UK infrastructure supporting leading research in the following areas:

Research on model organisms

The MRC is strengthening research using a number of organisms to model human disease, but with the mouse being the primary focus as it is generally agreed to be the most useful model. The MRC's priority is to boost research on mouse models by improving UK infrastructure (housing, transgenics facilities, embryo banking) for the whole of the research community. MRC are working with Celera to allow UK-based scientists to access proprietary mouse DNA sequences and clone sets that are not available publicly.

Macromolecular structure

Detailed knowledge of macromolecular structures is often needed before function can be understood or new therapies developed. MRC has set up facilities to improve methods of high throughput protein purification, crystallisation and analysis so that the UK biomedical research community can fully exploit the opportunities for more widespread structural analysis.

Theme 2: Translational Research on Major Clinical Problems

The MRC is strengthening clinical and translational research in key disease areas, using both grant funding and developing new centres, resources, and networks proactively:

Mental Health Research Priority

MRC is strengthening a whole range of research relevant to mental health, from basic studies on psychiatric genetics and biological mechanisms, imaging and neuroinformatics, through to work on effective health service delivery, in partnership with the Department of Health and the NHS.

Cardiovascular Research Priority

MRC, in partnership with the British Heart Foundation, are setting up centre(s) of excellence in cardiovascular disease research around individuals of international stature and having strong links to other key groups in the UK and USA.

Cancer Research Priority

MRC is developing its cancer portfolio in close co-operation with other funders of cancer research. Priorities include the development of resources needed to explore the genetic basis of cancer, cancer cell biology and prostate cancer research.

Human Population Based Studies

MRC has highlighted the need for large collections of well characterised human DNA samples for research to identify genes involved in the development or progress of disease and those affecting the response to treatment. Following a call for proposals early in 2000, MRC has funded a number of DNA collections from patient cohorts and case-control studies, covering a wide range of major causes of morbidity and death.

MRC, in partnership with the Wellcome Trust and the Department of Health, has set up a large prospective adult cohort to study the interactions between genetic and environmental risk factors for common multi-factorial diseases in a cohort of half a million people, to help determine the relative contributions of environmental and genetic factors to the development of chronic diseases (diabetes, stroke, heart attacks) in middle and later life.

New Medical Technologies

Advances in medical understanding will often need to be matched with new technologies to find application. MRC will continue to fund novel research into genetic screening, tissue regeneration/engineering (including stem cell therapy), immunotherapy, antibody engineering, surgery, gene therapy, and drug design – in addition to evaluations of existing and alternative treatments.

Innovative health technologies and the public

MRC will continue working with the Economic and Social Research Council (ESRC) to explore the factors that determine society's ability to benefit from advances in medical knowledge. The MRC is working closely with other Research Councils in co-ordinating efforts in areas such as technology development, bioinformatics, gene function research, and economic and social research.

4.1.5 Defra & FSA Genomics Programmes

A large proportion of Defra's genomics programmes focus on the genetic improvement of farm animals, poultry, fish and plants (fruit, vegetables, cereals and forage plants) including

the development of disease resistant varieties. Much research effort is also focussed on the molecular characterisation of animal and plant disease-causing pathogens and the development of molecular diagnostics to detect their presence. Over the last few years there has been an increasing concern in the UK over the environmental safety of GM crops. Several Defra programmes are currently running which seek to monitor the gene flow from GM to non-GM equivalent crops and to wild crop relatives. Finally, there are several Defra funded molecular ecotoxicology projects currently running which are investigating the effects of xenobiotics (including endocrine disrupting compounds) on aquatic and terrestrial invertebrates and vertebrates. The use of genomics in FSA funded research programmes falls broadly into two main categories: food authenticity and labelling; and food safety. FSA funded projects falling into the former category include the identification and quantification of meat and fish species in foods, the identification of meat species in vegetarian foods and the detection of olive oil adulteration with nut oils. In the food safety category, biomarkers are being developed and DNA microarrays used to study the effect of food chemicals, xenobiotics and toxins on apoptosis, cell proliferation and carcinogensis. The FSA also has several projects running which are assessing the safety of genetically modified foods. These include a project using transciptome, proteome and metabolome analysis to detect unintended effects of genetically modified potatoes, the development of methods for the analysis of GM wheat and barley seed for unexpected consequences of transgenic insertion and the development and comparison of molecular profiling methods for improved safety evaluation using GM brassicas.

4.2 International programmes

E.U. 5th and 6th Frameworks

5th Framework Projects

Title: New methodologies for assessing the potential of unintended effects in genetically modified (GM) food crops.

Project reference: QLK1-CT-1999-00765

Objective: To develop new methodologies that are of sufficient sensitivity and specificity to assess risks from this possible food borne hazard. Implicit in this objective is the need to develop new knowledge intended to understand the implications of the GM process on metabolic pathways in plants. Strong emphasis is placed on the exploitation of combinative and innovative 'cell factory' technologies (genomics, proteomics and metabolomics).

Coordinator: Dr Robert Bogers, State Institute for Quality Control of Agricultural Products, Agricultural Research Department, Wageningen, Netherlands.

Title: Development of systems to improve phytoremediation of metal contaminated soils through improved phytoaccumulation.
Project reference: QLK3-CT-2001-00429
Objective: To characterise metal-responsive genes using functional genomics in metal accumulating plant species and to isolate and modify and test endophytes and rhizosphere

microorganisms in combination with these plants. Risk assessment of the approach forms a central core of the studies. The goal is to develop systems for improved phyto-accumulation of metals.
Coordinator: Professor Matti Uusitupa, University of Kuopio, Kuopio, Finland.

Title: A functional genomic approach to measuring stress in fish aquaculture.
Project reference: Q5RS-CT-2001-02211
Objective: To develop whole animal gene expression profiles to identify candidate genes, which are associated with resistance to various stressors in rainbow trout. Information obtained from genomic expression profiles of stressed fish is a necessary base for further development of a marker-assisted selection strategy. The primary aim of this project is to utilise DNA microarray technology together with whole animal physiology to obtain gene expression profiles of fish tissues with and without stress, identify genes whose expression is linked to adaptations to stress and assess variations in gene expression in trout that show varying responses to stress to identify potential candidate markers of genotypic variation in stress responsiveness.
Coordinator: Dr Bernard Coquet, Centre de Recherche de Rennes, Institut National de la Recherche Agronmique Domaine de la Motte au Vicomte, Le Rheu, France.

6th Framework Proposals

Title: An animal free technology platform for identification of safe compounds.
Objective: To develop and implement an innovative validated technology platform facilitating the development of immunologically improved and safe therapeutic proteins, vaccines and chemical compounds. The project will utilise genomics, proteomics and bioinformatics to identify representative markers for allergenic, irritative or toxic characteristics of compounds into a high throughput assay format using state of the art nanotechnology.
Proposer: Novozymes A/S, Bagsvaerd, Denmark.

Title: Protozoa / Biomarkers / Biosensors / Genomics
Objective: To use ciliates as biomarkers/biosensors of heavy metals using genomics to identify molecular markers of stress responses to these metals.
Proposer: Dpto. Microbiologia-III. Facultad de Biologia, Universidad Complutense, Madrid, Spain.

Title: AQUAGENPROT
Objective: To use genomics and proteomics in cultivated fish and shellfish species as tools to improve safety, quality and benefits to human health from its consumption.
Proposer: Centre Tecnologico del Mar, Fundacion CETMAR, Vigo, Spain.

Title: Zebrafish Models for Human Development and Disease.
Objective: To combine the advantages of the zebrafish in mutagenesis screening and developmental biology with recent advances in functional genomics, in an integrated high-throughput project for the benefit of human health and basic research.
Proposer: Max-Planck-Gesellschaft zur Forderung der Wissenschaften e.V., Max-Planck-Insitut fur Entwicklungsbiologie Abteilung III – Genetics, Tubingen, Germany.

U.S.A.

Most of the toxicogenomics programmes currently funded in the U.S.A. are run through the U.S. Environmental Protection Agency, Oak Ridge National Laboratory and the National Center for Toxicogenomics (NIH). These programmes and their toxicogenomics research focus are listed below.

U.S. Environmental Protection Agency

Gulf Ecology Division and Mid-Continent Ecology Division

Use of molecular biomarkers to assess change in ecological status in aquatic ecosystems related to chemical contaminants & biotechnology products

Experimental Toxicology Division

Risk assessment for pulmonary, cardiovascular. Hepatic, renal and immunotoxicity at multiple levels – from molecular and biochemical through to whole animal models.

Reproductive Toxicology Division

Developing more sensitive methods, including molecular methods, to detect the adverse effect of environmental contaminants on human reproductive health

Environmental Carcinogenesis Division

Use of molecular techniques in cancer risk assessment pertinent to the mode of action of a chemical and the mechanisms by which it can induce tumours.

Molecular Ecology Branch of the National Exposure Research Laboratory

Development of genetic indicator methods including molecular finger-printing techniques (for analyzing genetic diversity), gene expression systems (for studying exposure to high priority chemical contaminants), cytogenetic and molecular methods for assessing exposure to genotoxic agents, and techniques for measuring DNA and protein adducts as molecular dosimeters in exposure monitoring studies.

National Centre for Environmental Assessment

Acts as an interface between researchers in the different components of the EPA and the regulators. Current initiatives include the organization (with the Office of Research and Development (EPA), National Institute of Environmental Sciences and American Chemical Society) of a conference "Developmental Toxicology in the 21st Century: Multidisciplinary Approaches to Using Model Organisms and Genomics" and the establishment of a Biotechnology Risk Assessment and Support programme which will evaluate current methodology, including molecular techniques, and where necessary develop new methods or new approaches to risk assessment of biotechnology products.

Oak Ridge National Laboratory

Risk and Regulatory Analysis Group, Life Sciences Division

Toxicology and risk analysis research endeavour spanning a broad spectrum from molecular toxicology to risk-based integrated assessments. Molecular research is focussed on genetic toxicology, human genome research and aquatic and environmental toxicology

Environmental Science Division Genomics Laboratory

Funded projects are currently centred around four bacteria of environmental importance: *Shewanella oneidensis* MR-1, *Deinococcus radiodurans* R1, *Rhodopseudomonas palustris*, *Nitrosomonas europaea*. The main focus of these projects is on the use of DNA microarray technology to elucidate gene function and regulation. Whole genome arrays of the above organisms are being constructed to profile gene expression under different environmental conditions.

National Center for Toxicogenomics (NCT), NIH.

NCT Microarray Group

The goals of the Microarray Group are:

- To identify toxicant-specific patterns of gene expression;
- To elucidate molecular mechanisms of environmental agents;
- To develop gene expression-based biomarkers of human exposure;
- To evaluate how well environmental effects of toxicants extrapolate from one species to another;
- To study the toxicological effects of chemical mixtures;
- To evaluate dose response curves and high or low dose-specific toxicological responses; and
- To develop a public database of microarray expression profiles.

Toxicogenomics Research Consortium

(NCT Micorarray Group with University of North Carolina at Chapel Hill, Fred Hutchinson Cancer Research Center, Oregon Health and Science University, Duke University and Massachusetts Institute of Technology)

The goals of the TRC are:

- To identify and characterize sources of variation in gene expression experiments and establish standard protocols ("best practices") for gene expression experiments and bioinformatics standards and tools;

- To evaluate toxicant-specific patterns of gene expression and define trans-species comparisons for responses to toxicant-induced stress;
- Integrate gene expression profiling data with other sources of data on proteomics, metabonomics and phenotypic anchoring;
- Elucidate molecular mechanisms of cellular responses to environmental agents;
- Develop gene expression-based biomarkers of human exposure;
- Study the toxicological effects of chemical mixtures;
- Evaluate toxicant dose-response behaviour and high/low dose-specific responses

Canada

There are three main centres conducting toxicogenomic research in Canada:

1. The Centre for Advanced Research in Environmental Genomics, University of Ottawa uses a broad range of molecular techniques to examine the interaction between genotypic variation and environmental heterogeneity including that brought about by environmental contaminants. The Centre is currently using a goldfish model to produce a hypothalamic microarray to explore the effects of exnooestrogens on hypothalamus function, this organ controlling the release of pituitary hormones regulating growth, reproduction and metabolism in both embryos and adults.

2. Caren Helbing's group, Department of Biochemistry and Microbiology, University of Victoria have designed a novel multispecies frog cDNA microarray to assess gene expression profiles of the model species Xenopus laevis exposed to endocrine-disrupting chemicals, including the herbicide acetochlor, that target thyroid hormone action.

3. The Molecular & Cellular Toxicology Lab, Great Lakes Institute, University of Windsor performs studies on molecular, cellular and physiological responses after organismal exposure to contaminants (in vivo, in vitro and field exposures). Catalytic enzyme, cellular, molecular and gene biological indicators (biomarkers) are used to define the cytochrome P450 activity and endocrine-disruption. Using hepatocytes perfused from fish species, laboratory validation of field toxicological studies are performed to assess the oestrogenicity and antioestrogenicity of selected and new emerging compounds such as PBDEs. Quantification at the gene level using molecular-gene assays, and by direct measurement of the expression of oestrogen-receptor and Ah-receptor mediated responses are investigated. Special focus is placed on differentially regulated genes after contaminants exposure.

Japan

The Ecotoxigenomics group in the Centre for Integrative Bioscience, National Institute for Basic Biology are using differential display and cDNA microarray methods to detect the

variation in oestrogen responsive genes to oestrogenic chemicals in *C. elegans*, Daphnids, Xenopus laevis, Medaka, mice and human umbilical cords. The group has also sequenced several thousand genes from cormorants showing high dioxin/PCBs exposure in Japan.

A group led by Masaharu Seno in the Department of Bioscience & Biotechnology, Okayama University Graduate School is using genes expressed in mouse mammary gland in an attempt to analyse time course dependent gene expression patterns for the assessment of life-long chemical exposure risk. Gene expression profiles of 22 genes have been examined with respect to time course and hormonal condition. A microarray of these genes will then be constructed and applied to the risk assessment of long term exposure to chemicals.

5. POTENTIAL OF ENVIRONMENTAL GENOMICS

5.1 What can it deliver to the Environment Agency?

5.1.1 Synopsis

The scientific benefits and potential of genomics to increase our understanding of the biology of organisms is immense. It is essential, therefore, that the impact of developments in genomics on the Agency's business is evaluated and the time frame that this impact will be seen in the business is established. Furthermore, by proactively assessing outputs from genomics research, we can add value to our current biological tests and methodologies. This will allow the Agency to examine the obstacles preventing their integration into the regulatory framework. Finally, the Agency should be aware of the context in which take-up of these tools will provide the greatest return and prioritize these areas for development.

To assist the Agency with this evaluation a table of its major science areas and relevant milestones is presented (Table 5.1) together with an objective analysis of the potential for genomics to deliver in these areas. A summary of the salient points arising from this analysis is provided below.

5.1.2 Genomics as an Information Provider

By far-and-away the greatest attribute of environmental genomics is its power as an information provider. The challenge faced, however, is the robust interpretation of the data to generate finite answers to specific questions because, the quantity of data provided is so large.

Notwithstanding this qualification, genomics is capable of providing:

- A better understanding of ***mechanisms of action*** – due to the level of molecular detail that is revealed (e.g. toxicological profiling of lung responses to diesel fumes). Genomics research could enhance our ability to identify the specific molecular targets of stressors and help us understand the effects of exposure to multiple stressors and stressors at low levels. It has significant potential when dealing with substances where traditional toxic endpoints are not the most relevant i.e. endocrine disruptors and general human pharmaceuticals. Testing based on gene regulation and protein expression is likely to challenge the traditional ecotoxicological test end-points (e.g. lethality).

- A more robust ***extrapolation of laboratory data to natural populations*** (as is necessary in risk assessment) will be possible as we improve our understanding of mechanisms of action.

- ***Predictive toxicity*** – the level of sensitivity inherent in measuring molecular parameters provides multiple sub-lethal indices. By mechanistically linking these to demographic or population level responses it will be possible to predict when low-level toxicity is occurring before it causes irreversible damage at the level of the individual or population.

This will enhance the "preventative" mode of environmental protection and diminish the necessity for "cure".

- Insights into ***sensitive sub-populations and species*** – through the identification of the molecular targets of pollution it may be possible to predict intra-species or population variability. This has particular significance in the context of the Habitats Directive when assessment is required of the impact of stressors on biodiversity where toxicity information is only available for a fraction of the biota.

- ***Understanding the sustainability of populations*** by, for example, providing data on gene flow. Wildlife populations are based on highly variable census information but this doesn't inform us about underlying population sustainability. "Omic" technologies will allow us to answer the more relevant questions about populations and communities.

- Confirmation that ***levels of protection are adequate*** in, for example, sensitive "windows" in an organism's development or in growth cycles (i.e. endocrine disruptor compounds are known to cause impacts at crucial stages of fish development).

- Potential to ***reduce uncertainties in risk assessment*** and facilitate a more rapid evaluation of a chemical's toxic potential i.e. new chemicals could be screened in parallel with known toxicants for similarity in expression profiles. This could be particularly useful for potentially genotoxic substances. It will challenge current definitions that separate carcinogens from non-carcinogens.

- An insight into ***genetic risk factors for disease*** (predisposition) and interactions between genetic factors and environmental exposure. These factors are being driven primarily for human populations, but the same approaches could be important for studying wildlife populations at risk (Habitats Directive/ Part IIA Contaminated Land Regulations).

5.1.3 Tools and techniques

Genomic advancement has been question-led but ultimately technology driven. The consequence of this is that the available genomic tool-box is evolving rapidly through the delivery of specific knowledge (i.e. biomarkers for defined biological endpoints) but also through the mechanical instrumentation used to derive this information. For example, the quantification of genetic end-point previously took 1-2 days but genetic assays recently developed to detect the SARS virus takes less than 4 hours (http://www.emory.edu/WHSC/HSNEWS/releases/apr03/sars_test.html). Due to the rapid nature of the platform technologies driving genomics, this summary will focus on the generic areas where new tools and techniques are becoming available. These include:

- ***Toxicity profiles*** – the measurement of the global pattern of gene expression, protein expression or metabolite content may generate a fingerprint for pollutant(s) providing identity, chemical class and biological effect.

- Identification and validation of ***novel targets and biomarkers*** of exposure (e.g. vitellogenin in male fish exposed to endocrine disruptor compounds and quantification of bioavailability of cadmium to earthworms through metallothionein measurement).

- Provision of viable ***alternatives to whole animal testing*** – ability to match test organisms to particular environmental questions. There is the potential for rapid turn around of information and for making stressor-specific tests portable.

- Specification of ***data frameworks for ecotoxicology*** – due to the sensitivity of molecular indices it has been necessary to define very stringently the parameters under which ecotoxicological experiments are performed. Furthermore, the complexity of the data sets necessitates the accurate and comprehensive recording of all relevant data. These schemes form the basis of relational data structures which allow advance data mining.

- Development of ***objective informatic frameworks for risk assessment –*** the relational nature of genomics information underpins the construction of predictive models of toxicity.

5.1.4 Obstacles, data gaps and solutions

It is important to manage the expectation for genomics to deliver as there are significant hurdles which must be overcome before its full potential can be realized. These include.

- ***Lack of genomic knowledge of non-model organisms*** – there is a significant gap in knowledge relating to genetic information about 'non-model organisms' (those not routinely used in ecotoxicological experiments). As the genome sequences of more 'model' organisms are completed (e.g. human, mouse, *Drosphila*, *C. elegans* and yeast), the science of ecological risk assessment will have to consider how to deal pragmatically with filling this knowledge gap.

- ***Robustness of the data generated*** - ongoing research programmes will need to deliver high standards of data before regulatory decisions can be made based on their findings. The same issue of confounding factors [e.g. age, sex and environmental stimuli other than the stressor(s)] exists within measured changes in gene expression. Experimental models need to be well validated before firm conclusions can be drawn. These issues have been recognized through the requirement of exacting experimental descriptions when generating "omics" data.

- ***Limited validated data*** – The naivety of this area means that there are few comprehensive case studies. It is important to accelerate data into the public domain to facilitate shared learning.

- ***Lack of baseline data*** – research leading to baseline and confirmatory data (through ecological and chemical monitoring programmes already being conducted by the Agency and others) must be continued to enable the benefit from the 'omic techniques to be taken up within the Agency.

- ***Uncertainty -*** The harm debate will be reinvigorated. Is an impact at the gene level considered to be 'harm'? In the short to medium-term there is likely to be an increase (rather than a decrease) in uncertainty in ecological risk assessment. There are possible health-risk communication issues surrounding perception of gene damage, particularly in humans.

Table 5.1 Genomics Delivery for Environment Agency Science Areas

Science Area	Relevant milestones	How can genomics help?	What are the obstacles	Priority setting	Timescales
Air Quality	What are the relative impacts of air pollutants from Agency regulated processes including accidental releases (COMAH)?	The toxicological impact of air pollutants on lungs is well-established including modes of action and reversible and irreversible effects. Value added information is now being provided by genomics and this will provide additional power for the interpretation and analysis of chronic and acute pollutants (Section 3.5). Furthermore, subsequent to validation, *In vitro* tools may replace the present animal models.	These rely on animal models but R&D is currently underway to develop in vitro tests using genomics. Data is scarce on non-vertebrate impacts of air borne pollutants. Roadside and urban ecosystems have not been studied using genomics and an investigation of the potential for food web transmission should be a priority.	High	Blue Sky – Complete Academic Acceptance - established for in vivo, required for *In vitro*. International Standardisation - 3-5 Years.
Climate Change	What are the likely direct and indirect impacts of climate change on water and air quality, land capability, water resources and demand, flood risk, human health, biodiversity, fisheries, waste disposal and nuclear decommissioning?	Two distinct areas where genomics can contribute: Laboratory modeling and analysis can evaluate the sub-lethal impact of high temperature and high CO_2 on a variety of organisms (transgenic modeling of such effects is underway; Harwood 2003); population profiling (microbial and eucalypti) can be used to evaluate genetic diversity. Continuous monitoring will provide a better understanding of the likely impacts of climate change.	The effect of mitigating factors such as normal climate variation and population change acting over geological timescales on genomic end points is hard to predict. Significant R&D will be required before significant confidence can be attributed to the data generated.	Low	Blue Sky - Underway Academic Acceptance - 5-10 years International Standardisation – >10 years

Data management and modeling	How can monitoring information be collated and manipulated to answer new questions (particularly for land, air, water, biodiversity, natural resources and waste and climate change trends)? How can we improve the use of existing data and models to understand environmental processes, target activities and reduce our requirements for additional monitoring or management information? How can we improve our use and statistical manipulation of data to reduce risk and uncertainty across all activities? How do we best influence the external modeling community and what are the best long-term strategies for model development and support?	Genomics can provide data to objectively measure biological impact. If interpreted and annotated correctly the diagnostic power of genomics will inform and allow predictive models to be improved.	At present, collection of appropriate basal level information is not being performed. By including appropriate genomic end points in surveys such as Countryside Survey 2000 (CS-2000) (2006) the benefits of the additional information provided by genomics may be realized. Furthermore, appropriate computational infrastructure (hardware and software) is required to facilitate the relational interpretation of sample meta-data. Through concepts such as the GRID it should be possible to couple genomic information with data such as geochemical and meteorological data originating from remote/unrelated sources.	Medium	Blue Sky - Underway Academic Acceptance - 4 Years. International Standardisation - 10-20 years
Diffuse pollution	What are the relative impacts on water quality of point source and diffuse pollution arising from both urban and rural areas? What is the contribution of contaminated sediment and silt to poor water quality and ecological status and how can it best be managed?	The development of predictive markers to monitor biological impact of diffuse pollution will allow remedial action to be initiated prior to irreversible damage being observed.	Baseline data requirement particularly in relation to relative role of sediment to poor water quality in relation to other pressures.	Medium	Blue Sky

Ecological effects	What are the effects of specific chemicals, pathogens and nutrients on ecological systems? What are the effects of complex emissions, multiple inputs and long-term exposure to low doses on ecological systems? What other activities have impacts on environmental quality and how can the Agency best manage, or influence others to manage, these?	The development of predictive markers to monitor the biological impact of specific chemicals and pathogens will allow remedial action (or non-release) to be initiated prior to irreversible damage occurring. Genomics has the ability to integrate the effects of complex emissions to provide a single biological measure, which can then predict long-term damage of exposure to low doses. The development of the VTG assay as a sub-lethal marker for fish feminisation (Section 3.3) is the best development relevant marker.	The requirement to measure environmental damage and therefore to define natural variation. The reporting of genomic change, which has no biological/ecological relevance. The need to establish mechanistic links to irreversible effects. The levels at which populations may collapse needs to be firmly established. Furthermore, the relevance to niche species and organisms other than fish (such as invertebrates) requires substantial research.	High	Blue Sky: Principal established. Academic Acceptance: Each endpoint requires validation; some are established, others will be forthcoming over the next 10 years. International Standardisation: Some ring tests are completed some are underway. Legislation with 1-2 years. Understanding endocrine and other sub lethal effects in invertebrates is a very underdeveloped area and is the subject of a DEFRA initiative. Academic Acceptance: VTG assay well accepted. International Standardisation: Some ring tests are completed, some are underway. Legislation within 1-2 years.
Fisheries	What are the remaining pressures on fish stocks (including food availability, predation, disease and parasites, water quality, change in flow and physical alterations) and	Population modeling can be used to determine the genomic heterogeneity of populations and to determine the sustainability of a population. Dietary health, reproductive success and disease status can be determined	Little or no base line data exists.	Low	Blue Sky: Majority. Academic Acceptance: Population testing has been used for fish population studies and data used in legal cases.

	how can these best be managed?	using gene profiling. Genomics can, therefore, provide 'value added' information.			International Standardisation: None
Forecasting and decision analysis	How can scenarios and 'what if' questions be best used to provide a reasoned basis for considering future changes in environmental pressures or impacts and hence improve decision analysis and environmental policy formation?	Full genomic analysis will provide a large body of unconstrained information from which to better inform scenario building. Short-term laboratory models can be used to develop predictive models, which can then inform policy.	Interpretation of complex data sets. NOECs will not exist due to the sensitivity of transcriptomics. Therefore, employing the precautionary principal solely on the basis of transcriptomic data will provide inflated risk models.	Medium	Blue Sky: Majority Academic Acceptance: 3 years International Standardization: 10 years.
Human health effects	What are the effects of specific chemicals and pathogens on human health? What are the effects of complex emissions, multiple inputs and long-term exposure to low doses on human health?	Subject area - toxicogenomics. Human toxicology is being substantially empowered through genomics. Predictions of impact on human health are already in use, employing animal models and cell cultures within pharmaceutical testing regimes.	Toxicological interpretation in some cases still needs to be established and the relevance of some model organisms and cell cultures is yet to be completed.	High	Blue Sky: Complete Academic Acceptance: Majority International Standardisation – 5-10 years.
Monitoring strategy and tools	What are the most appropriate indicators for land and soil quality, groundwater quality, air, natural resources, waste, biodiversity, flood and asset risk and climate change trends?	By providing a tiered series of endpoints that can provide generic information on integrated toxic impact to the exposure of a specific bioavailable toxic component such as cadmium.	Long-term studies are establishing marker efficacy and long-term population level relevance.	Medium	Principal established Academic Acceptance: A small number. International Standardization: 5-10 years.
Resource and habitat requirements	What is an appropriate level of ecological quality to protect and how do we measure and manage this under conditions of natural variability?	Determination of the genomic complexity of a population can be used to model the resilience of the population. In turn genetic diversity can give information on whether the population can cope with additional stress. These techniques have already	Baseline data is required for natural viability of genomic parameters. Appropriate tools do not exist for many species and bespoke development is required.	Low	Blue Sky: Principal established Academic Acceptance: Established for a small number of species. International standardization:

	How can we measure the resilience of specific species to different types of stress? What are the impacts of invasive species and disease on environmental management activities for which we are responsible for or can influence?	been used to inform the cross-breeding and transplantation of endangered species.			Ecosystem specific tools always required.
Risk assessment	How can we meaningfully rank the relative importance of different environmental risks such as global warming, acid deposition and flooding taking into account both scientific and public perception criteria – what is acceptable risk?	Provision of additional information.	Baseline data requirement	Low	Blue Sky: Majority
Soil protection and land management	What are the impacts of different types of rural and urban land use on soil quality, water resources, flood risk, quality of life and biodiversity and how can these best be managed? What are the most appropriate technologies for dealing with contaminated land and landfill sites and how can we restore appropriate soil function for different types of uses?	Microbial invertebrate population profiling can be used to evaluate these situations (Section 3.1).	Interpretation of data with regards the impacts of natural variation.	Medium	Blue Sky: Principal established Academic Acceptance: Established for in small number of contexts. International standardization: 5-10 years.

- ***Requirement for multidisciplinary teams*** – a concerted effort is required of scientists across a broad span of disciplines to help in the interpretation of the myriad data that will be produced by the "omic" technologies. This will require sophisticated pattern recognition techniques and bioinformatic approaches.

5.2 Implications of Genomics for the end-users and society

5.2.1 Introduction

Genomics will have a significant influence on industry and regulators as key end-users. The power of genomics, correctly employed, should lead to more environmentally-friendly products and cleaner processes, and therefore contribute to sustainable development. However, the threat comes from the uninformed or incorrect implementation/interpretation of genomic data which could lead to an overestimation of risk which, in turn, could lead to unnecessary public concern and industrial restriction. The consequence of this could be that genomic analysis leads to an increase in uncertainty in the short-term. However, if correctly managed, this uncertainty should diminish as our knowledge and ability to interpret the data increases. A case study illustrating this specific issue is provided in Section 3.1. Several key issues are identified below where genomics will impinge on the Agency and that might requires some Agency activity in response.

5.2.2 Building end user communities

There are a number of initiatives that aim to bring together the relevant end users and scientists together, including NERC Environmental Genomics Community Programme workshops, SETAC International Ecotoxicogenomics workshop (Florida 2002) and International Council of Chemical Associations (ICCA) Workshop (March 2001, Orlando, Florida). Whilst the Agency has attended some of these, it will be important to steer the Agency and its stakeholders to discussions on the aspects of genomics relevant to its regulatory goals. Moreover, the Agency has been proactive in generating this review, yet other regulatory end users may still be unaware of the potential risks and benefits to their regulatory areas. A workshop or workshops to take forward these agendas would be a practical approach.

The ICCA Workshop brought together leading academics, representatives from industry and regulators to identify the core issues and benefits that genomics could offer to (eco)toxicology and epidemiology. The wider ethical and social implications of 'omic technologies was also discussed. The benefits and issues were raised regarding the application of 'omic technologies to (eco)toxicology are summarized below:

Top five benefits for the application of 'omic technologies to (eco)toxicology?

- Improved mechanistic understanding (mode of action; MOA).
- Improved inter-species extrapolation (comparative genomics).

- Predictive toxicology and the increased application of information technology (IT) to safety assessment.
- Identification of sensitive sub-populations (intra-species variability).
- Identification of novel targets and biomarkers.

Top five issues regarding the application of 'omic technologies to (eco)toxicology?

- The need for examples in the public domain; shared learning (0-2 years).
- Clarity of the regulatory position (potentially referable findings) - molecular adaptation does not necessarily equate to toxicity.
- Provision of a centralized database.
- Evaluation of the predictive value of omics data.
- Genome closure and improved annotation especially for wildlife populations.

Recommendation for a strategy for the chemical and pharmaceutical industries to address these issues?

- Generate case studies and support research projects that evaluate how "omics" can improve risk assessment by studying well known compounds.
- Track and partner with key researchers.
- Engage in a database discussion to ensure the informatics infrastructure is being developed to meet the industry need.
- Form an international consortium among all sectors to provide guidance.

5.2.3 Impact on Pesticide Registration

One area in which genomics data may have an immediate impact, and will require discussion between regulators and industry, is within the "potential referable findings" clause of the pesticide registration framework. At present all relevant data must be presented to the regulators (in this case Pesticides Safety Directorate) when applying for registration. As the toxicological relevance of many genomic endpoints is, as yet, undefined, all such information must be presented. This will present regulators with a potential problem as chemicals may initiate transcript level changes which either have little biological relevance or whose biological relevance is as yet undefined. If release is granted and subsequently an effect is recorded that can be linked to the transcript data presented, the liability may fall on the regulators. However, if release is not granted on the basis of a transcript change with no established endpoint or even if additional toxicological data is requested, this may deny the public the benefits of a valuable product and impact the financial competitiveness of industries presenting transcriptomic data.

5.2.4 International standardization

The issue relating to the standardization of microarray experiments was recognized in November 1999 with the formation of the Microarray Gene Expression Data (MGED) Society. MGED is an international organization of biologists, computer scientists, and data analysts that aims to facilitate the sharing of microarray data generated by functional genomics and proteomics experiments. The focus is on establishing standards for microarray data annotation and exchange, facilitating the creation of microarray databases and related software implementing these standards and promoting the sharing of high quality, well-annotated data within the life-sciences community. The Society has already published the Minimum Information About a Microarray Experiment (MIAME, http://www.mged.org/Workgroups/MIAME/miame.html) document which outlines the minimum information that should be reported about a microarray experiment to enable its unambiguous interpretation and reproduction. This scheme has been adopted by major scientific journals who now require this information prior to publication.

Working groups are presently refining standards for toxicology (MIAME-Tox, http://www.mged.org), proteomics (HUPO, http://psidev.sourceforge.net/meetings/2002-10/report.htm) and ecotoxicology (MIAME-Ecotox - http://envgen.nox.ac.uk/). These standards provide guidelines for the user community when they are exploiting data whilst also supplying a level of confidence in genomic data when it is being presented and evaluated.

6. CONCLUSIONS AND RECOMMENDATIONS

6.1 Concluding remarks

It is very timely for the Agency to be addressing the subject of genomics. The academic and industrial skills-base is developing in the UK with substantial infrastructure and fundamental research being supported, mainly through the Office of Science and Technology's genomics initiative. It is essential that the Agency develops its own strategy for genomics so that it can be proactive and take advantage of the opportunities presented. In addition, by being proactive the Agency will be in a position to protect itself against the potentially controversial issues that may arise through the use of genomics in an environmental context, or contrast with the Agency's duties.

Academic, government and industrial scientists have been investigating the biological, chemical and physical impacts of toxicants on gene and protein expression in wildlife populations for a number of years. Perhaps the most widely studied is the expression of the protein vitellogenin in fish exposed to oestrogenic stimuli. Rather than studying single gene or single proteins that respond to environmental stimuli, the advent of tools to examine 'omic-level responses has increased the scale and complexity with which these measurements can be made. The sheer number of potential opportunities requires the Agency to identify where genomics can have an immediate impact within its science strategy and prioritise accordingly. This strategy should include the underpinning of the Water Framework Directive where it can provide additional information relating to ecological structure and function. However, it is also essential to moderate expectation and realize the limitations and potential threats posed by genomic technologies.

Environmental genomics has the potential to act as the vehicle through which the public can be made aware of, and can understand the direct benefits of genomics. However, for this to be delivered it is essential that the Agency manages this issue carefully and proactively. With the potential for irresponsible presentation of non-validated data into the public domain, leading to unsubstantiated public concern, there is a requirement for genomics issues to be addressed promptly if risks are to be minimised. This report has striven to illustrate the phenomenal potential of genomics whilst identifying potential shortcomings and bottlenecks. The benefits and risks outlined provide the Agency with the basis from which to establish a policy on Environmental Genomics.

6.2 Recommendations

Specific actions recommended to the Agency include:

- ***To become more aware*** of the potential of 'omic tools, but also to highlight the potential for misinterpretation, especially in the short term while rigorous validation is underway.

- ***Identify where the technology can be employed*** by the business in the short to medium-term. Prioritize where it could have an immediate impact – likely Endocrine Disruption and Water Framework Directive where underpinning information relating to ecological

structure and function could be provided. Begin the necessary preparations which will enable its uptake – scientific and policy challenges.

- ***A rapid mechanism for validation and adoption of new tools***. Flexible arrangements are needed to take advantage of opportunities as they arise, such as the validation of novel biomarkers so they can reduce, refine and replace whole organism tests in the long-term.

- ***Address the Agency's knowledge base*** considering the potential issues associated with interpretation and assessment of genomic data i.e. technical competency requirements of Agency staff.

- ***Collaborate nationally and internationally with genomic researchers*** to ensure they understand the issues faced by regulators i.e. legislative requirements, expectations of stakeholders (the human health agenda), to ensure research effort is focused on 'real issues'. Conversely, whilst there still remains significant pure academic science to be done, it is important to spread risk and investment by working with other funders such as NERC and Defra to deliver the technology validation required.

- ***Generate high quality case studies*** and support research projects that evaluate how 'omics can improve risk assessment by studying well-known compounds that are of concern to regulators.

- ***Develop predictive risk assessment software*** that can interpret and integrate 'omics data with established toxicity measures to generate objective risk assessment. And,

- ***Increase dialogue with other regulatory organizations*** – discuss lessons learned, evaluate the potential for collaboration on programmes of research, increase consistency in response to genomic data delivered to regulators (i.e. SEPA, PSD, VMD) and stilmulate the building of an informed end-user community.

In conclusion, these tools will increase our understanding of mechanisms of action, hazard identification, dose response relationships, interspecies variability and sustainability. The eventual outcome could be that regulatory bodies are able to identify 'polluting industries' that cause ecological impacts via specific conserved modes of action. But genomics will only be useful if the right data are collected, that the data can be interpreted and the implications understood, and then used effectively to inform policy decisions. This will require the Agency to stake steps now towards this strategic ideal if the benefits are to be achieved.

APPENDIX A GLOSSARY OF TERMS

Amplification: an increase in the number of copies of a specific DNA fragment.

Annealing: see DNA Annealing

Annetocin: A neuropeptide hormone with high homology to members of the vasopressin/oxytocin superfamily of neurohypopysial hormones. Evidence suggests that annetocin elicits stereotypical egg-laying behaviours in some invertebrates including ovulation and oviposition.

Apoptosis: The most common form of physiological (as opposed to pathological) cell death. Apoptosis is an active process requiring metabolic activity by the dying cell. Often called programmed cell death, although this is not strictly accurate.

Bioinformatics: the science of informatics as applied to biological research. Informatics is the management and analysis of data using advanced computing techniques. Bioinformatics is particularly important as an adjunct to genomics research, because of the large amount of complex data this research generates.

Biomarker: observable change (not necessarily pathological) in the function of an organism, related to a specific exposure or event.

Chorion: Protective membrane around the eggs of insects and fishes.

Chromosome: The DNA in a cell is divided into structures called chromosomes. Chromosomes are large enough to be seen under a microscope. In humans, all cells other than germ cells usually contain 46 chromosomes: 22 pairs of autosomes and either a pair of X chromosomes (in females) or an X chromosome and a Y chromosome (in males). In each pair of chromosomes, one chromosome is inherited from an individual's father and one from his or her mother.

Clone: A term which is applied to genes, cells, or entire organisms which are derived from - and are genetically identical to - a single common ancestor gene, cell, or organism, respectively. Cloning of genes and cells to create many copies in the laboratory is a common procedure essential for biomedical research. Note that several processes which are commonly described as cell 'cloning' give rise to cells which are almost but not completely genetically identical to the ancestor cell. 'Cloning' of organisms from embryonic cells occurs naturally in nature (e.g. with the occurrence of identical twins). The laboratory cloning of a sheep ('Dolly') using the genetic material from a cell of an adult animal has recently been reported.

Cloning: the process of producing a genetically identical copy (clone).

Cloning vector: DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vectors capacity for self-replication; vectors introduce foreign DNA into host cells, where it can be reproduced in large quantities. Examples are plasmids, cosmids, and yeast artificial chromosomes; vectors are often recombinant molecules containing DNA sequences from several sources.

Coding regions: those parts of the DNA that contain the information needed to form proteins. Other parts of the DNA may have non-coding functions (e.g. start-stop, pointing or timer functions) or as yet unresolved functions or maybe even 'noise'.

Codon: a set of three nucleotide bases in a DNA or RNA sequence, which together code for a unique amino acid. For example, the set AUG (adenine, uracil, guanine) codes for the amino acid methionine.

Complementary DNA (cDNA): Viral reverse transcriptase can be used to synthesize DNA that is complementary to RNA (e.g. an isolated mRNA). The cDNA can be used, for example, as a probe to locate the gene or can be cloned in the double-stranded form.

Deletion: in the process of DNA replication, a deletion occurs if a nucleotide or series of nucleotides is not copied. Such deletions may be harmless, may result in disease, or may in rare cases be beneficial.

DNA (Deoxyribonucleic Acid): the molecule that encodes genetic information. DNA is a double-stranded helix held together by bonds between pairs of nucleotides.

DNA Adducts: Chemicals (alkylating agents) may form covalent bonds with deoxyribonucleic acid (DNA) to form an adduct in which a methyl or ethyl group is added. DNA adducts are believed to play a major role in mutagenesis and clastogenesis, as well as in carcinogenesis

DNA annealing: The sticking together (renaturing) of complementary single strands of DNA to make a double-stranded DNA after the strands have first been pulled apart (denatured). This involves forming hydrogen bonds between the base pairs.

DNA probe: a piece of single-stranded DNA, typically labelled so that it can be detected (for example, a radioactive or fluorescent label can be used), which can single out and bind with (and only with) another specific piece of DNA. DNA probes can be used to determine which sequences are present in a given length of DNA or which genes are present in a sample of DNA.

DNA repair genes: genes which code for proteins which correct 'mistakes' in DNA sequences. When these genes are altered, mutations may be able to accumulate in the genome, ultimately resulting in disease. See genetic mutation, p53 and suppressor gene.

DNA replication: the process of making copies of strands of DNA. Existing DNA is used as a template for synthesising the new strands.

Dot blot: Method for detecting a specific message. A spot of solution is dotted onto nitrocellulose paper, a specific radiolabeled complementary probe is allowed to bind.

Ecotype: A subspecies, or breed, that has adapted to its local environment and, as a result, is distinctive from other members of the species but can still successfully interbreed with other members of the species.

Electrophoresis: A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different

rate, depending on its electrical charge and size. Separation is based on these differences. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

Endocrine Disrupting Chemical: An exogenous substance that causes adverse health effects in an intact organism, or its progeny, consequent to changes in endocrine function.

Endocrine glands: Groups of cells specialized to synthesize hormones and secrete them into the blood to regulate other types of cells. Examples are pituitary, thyroid, parathyroid, adrenal glands, ovary and testis, placenta and B cells of pancreas.

Endonuclease: An enzyme that cleaves its nucleic acid substrate at internal sites in the nucleotide sequence.

Eutrophic: Rich in dissolved nutrients, photosynthetically productive and often deficient in oxygen during warm weather (compare with oligotrophic).

Exogenous DNA: DNA which has been introduced into an organism but which originated outside that organism (e.g. material inserted into a cell by a virus).

Exon: exons are those portions of a gene which code for proteins.

Exonuclease: An enzyme that digests the ends of a piece of DNA.

Expressed sequence tag (EST): a short strand of DNA (approximately 200 base pairs long) which is part of a cDNA. Because an EST is usually unique to a particular cDNA, and because cDNAs correspond to a particular gene in the genome, ESTs can be used to help identify unknown genes and to map their position in the genome.

Full gene sequence: the complete order of bases in a gene. This order determines which protein a gene will produce.

Gene: a length of DNA which codes for a particular protein, or in certain cases a functional or structural RNA molecule.

Gene expression: The process by which a gene's coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Gene Families: Groups of closely related genes that make similar products.

Gene Library: A collection of cloned DNA fragments which, taken together, represent the entire genome of a specific organism. Such libraries or 'gene banks' are assembled so as to allow the isolation and study of individual genes. Gene libraries are produced by first breaking up or 'fractionating' an entire genome. This fractionation can be accomplished either by physical methods or by use of restriction enzymes. The genome fragments are then cloned (multiplied in number) and stored for later use.

Gene Product: the protein produced by a gene.

Gene sequence: The order of nucleotide bases in a DNA molecule that constitute a gene.

Genetic Code: the set of codons in DNA or mRNA. Each codon is made up of three nucleotides which call for a unique amino acid. For example, the set AUG (adenine, uracil, guanine) calls for the amino acid methionine. The sequence of codons along an mRNA molecule specifies the sequence of amino acids in a particular protein.

Genetic engineering (gene manipulation, genetic manipulation): The manipulation of an organism's genetic endowment by introducing or eliminating specific genes through modern molecular biology techniques. A broad definition of genetic engineering also includes selective breeding and other means of artificial selection.

Genetic Map: a map of a genome which shows the relative positions of the genes and/or markers on the chromosomes.

Genetic Mutation: a change in the nucleotide sequence of a DNA molecule. Genetic mutations are a kind of genetic polymorphism. The term 'mutation', as opposed to 'polymorphism', is generally used to refer to changes in DNA sequence which are not present in most individuals of a species and either have been associated with disease (or risk of disease) or have resulted from damage inflicted by external agents (such as viruses or radiation).

Genetic Polymorphism: a difference in DNA sequence among individuals, groups, or populations (e.g. a genetic polymorphism might give rise to blue eyes versus brown eyes, or straight hair versus curly hair). Genetic polymorphisms may be the result of chance processes, or may have been induced by external agents (such as viruses or radiation). If a difference in DNA sequence among individuals has been shown to be associated with disease, it will usually be called a genetic mutation. Changes in DNA sequence which have been confirmed to be caused by external agents are also generally called 'mutations' rather than 'polymorphisms'.

Genetic Predisposition: susceptibility to a disease which is related to a genetic mutation, which may or may not result in actual development of the disease.

Genetically Modified Organism: The modification of the genetic characteristics of a micro-organism, plant or animal by inserting a modified gene or a gene from another variety or species. Genetically modified organisms (GMOs) may be micro-organisms designed for use as biopesticides or seeds that have been altered genetically to give a plant better disease resistance or growth.

Genomic DNA: The basic chromosome set consisting of a species-specific number of linkage groups and the genes contained therein.

Genome: all the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

Genomic Library: A collection of clones made from a set of randomly generated overlapping DNA fragments representing the entire genome of an organism.

Genomics: the study of genes and their function. Recent advances in genomics are bringing about a revolution in our understanding of the molecular mechanisms of disease, including the

complex interplay of genetic and environmental factors. Genomics is also stimulating the discovery of breakthrough healthcare products by revealing thousands of new biological targets for the development of drugs, and by giving scientists innovative ways to design new drugs, vaccines and DNA diagnostics. Genomics-based therapeutics include 'traditional' small chemical drugs, protein drugs, and potentially gene therapy.

Genotoxin: A toxin (poisonous substance) which harms the body by damaging DNA molecules causing, for example, mutations or tumours.

Genotype: the particular genetic pattern seen in the DNA of an individual. 'Genotype' is usually used to refer to the particular pair of alleles that an individual possesses at a certain location in the genome. Compare this with phenotype.

GMO: see Genetically Modified Organism.

Housekeeping genes: The genes which are expressed in all cells and which code for molecules that are necessary for basic maintenance and essential cellular functions.

Heterologous Expression Systems: systems that allow expression of a gene in a different organism.

Human Genome Project: an international research effort aimed at discovering the full sequence of bases in the human genome. Led in the United States by the National Institutes of Health and the Department of Energy.

Human Genome Initiative: Collective name for several projects begun in 1986 by DOE to (1) create an ordered set of DNA segments from known chromosomal locations, (2) develop new computational methods for analyzing genetic map and DNA sequence data, and (3) develop new techniques and instruments for detecting and analyzing DNA. This DOE initiative is now known as the Human Genome Program. The national effort, led by DOE and NIH, is known as the Human Genome Project

Hybridization: The process of joining two complementary strands of DNA or one each of DNA and RNA to form a double-stranded molecule.

In Situ Hybridization (ISH): Use of a DNA or RNA probe to detect the presence of the complementary DNA sequence in cloned bacterial or cultured eukaryotic cells.

Intron: a length of DNA which is interspersed among the protein-coding sequences (exons) in a gene. Introns are transcribed (see transcription) into mRNA but are then cut out of the mRNA sequence before protein synthesis occurs.

Library: a set of clones of DNA sequences from an organism's genome. A particular library might include, for example, clones of all of the DNA sequences expressed in a certain kind of cell, or in a certain organ of the body.

Macroarray: A low-density array of DNA molecules used for parallel hybridisation analysis (see microarray).

Marker: a sequence of bases at a unique physical location in the genome, which varies sufficiently between individuals that its pattern of inheritance can be tracked through families

and/or it can be used to distinguish among cell types. A marker may or may not be part of a gene. Markers are essential for use in linkage studies and genetic maps to help scientists to narrow down the possible location of new genes, and to discover the associations between genetic mutations and disease.

Messenger RNA (mRNA): the DNA of a gene is transcribed (see transcription) into mRNA molecules, which then serve as a template for the synthesis of proteins.

Metabonome/Metabolome: constituent metabolites in a biological sample.

Metabonomics/Metabolomics: techniques available to identify the presence and concentrations of metabolites in a biological sample.

Metallothioneins (MTs): Ubiquitous low molecular weight proteins and polypeptides of extremely high metal and sulfur content. They are thought to play roles both in the intracellular fixation of the essential trace elements zinc and copper, in controlling the concentrations of the free ions of these elements, in regulating their flow to their cellular destinations, in neutralising the harmful influences of exposure to toxic elements such as cadmium and mercury and in the protection from a variety of stress conditions

Microarray: a component of a device for screening genomic or cDNA for mutations, polymorphisms or gene expression. The array is a small glass slide or other solid surface on which thousands of immobilized oligodeoxynucleotide probes have been synthesized or robotically deposited in a predetermined array, so that automated recording of fluorescence from each of the spots may score successful hybridizations. An array may be designed for the detection of all known genes of a species or selected specific sequences. The array may also consist of different antibodies or proteins.

Mutation: A change, deletion, or rearrangement in the DNA sequence that may lead to the synthesis of an altered inactive protein the loss of the ability to produce the protein. If a mutation occurs in a germ cell, then it is a heritable change in that it can be transmitted from generation to generation. Mutations may also be in somatic cells and are not heritable in the traditional sense of the word, but are transmitted to all daughter cells.

Nematode Maturity Index: Nematode assemblages are a potential measure of the quality of submersed, temporarily submerged, and terrestrial soils, and for the development of an ecological typology and biomonitoring system. The maturity index uses a colonizer/persister scale for nematodes to provide an assessment of the condition of the soil ecosystem.

Nitrocellulose membrane: A membrane with a high nonspecific absorbing power for biological macromolecules. Very important as a receptor in blot-transfer methods.

Non-genotoxic Carcinogen: a substance that causes cancer, not by primarily damaging the genetic material, but by mechanisms that stimulate cell proliferation, thus increasing the chances for natural mutations to be reproduced, and/or selection of specific cell populations that may derange in a later stage.

Northern blot: An electroblotting method in which RNA is transferred to a filter and detected by hybridization to32 P-labelled RNA or DNA.

Nucleic Acid: one of the family of molecules which includes the DNA and RNA molecules. Nucleic acids were so named because they were originally discovered within the nucleus of cells, but they have since been found to exist outside the nucleus as well.

Nucleotide (= base): the 'building block' of nucleic acids, such as the DNA molecule. A nucleotide consists of one of four bases - adenine, guanine, cytosine, or thymine - attached to a phosphate-sugar group. In DNA the sugar group is deoxyribose, while in RNA (a DNA-related molecule which helps to translate genetic information into proteins), the sugar group is ribose, and the base uracil substitutes for thymine. Each group of three nucleotides in a gene is known as a codon. A nucleic acid is a long chain of nucleotides joined together, and therefore is sometimes referred to as a 'polynucleotide'.

Nucleus: the membrane bound structure containing a cell's central DNA found within all eukaryotic cells.

Oestradiol (= estradiol (USA); follicular hormone): A hormone synthesized mainly in the ovary, but also in the placenta, testis, and possibly adrenal cortex. A potent oestrogen.

Oestrogen (= estrogen (USA): A steroid sex hormone that regulates female reproductive processes and creates feminine secondary sexual characteristics. Some types of cancer depend on oestrogen for their growth, and modern-day increases in cancer rates are thought by many to be caused by the action of certain chlorinated organic chemicals (such as pesticides and herbicides) that mimic the action of this hormone.

Oligonucleotide: A molecule made up of a small number of nucleotides, typically fewer than 25. These are frequently used as DNA synthesis primers.

Oligotrophic: Describes a body of water in which nutrients are in low supply.

Oncogene: a gene which is associated with the development of cancer.

Organelle: Sub-cellular structures that perform a role within each cell. These vary widely from the nucleus, containing all our genetic information, to the golgi apparatus which processes protein and secretes it in vesicles on demand by the cell.

Oxytocin: A peptide hormone from hypothalamus: transported to the posterior lobe of the pituitary. Induces smooth muscle contraction in uterus and mammary glands.

PCR: see Polymerase Chain Reaction.

Pharmacogenomics: The science of understanding the correlation between an individual patient's genetic make-up (genotype) and their response to drug treatment. Some drugs work well in some patient populations and not as well in others. Studying the genetic basis of patient response to therapeutics allows drug developers to more effectively design therapeutic treatments.

Phenotype: a set of observable physical characteristics of an individual organism. A single characteristic can be referred to as a 'trait', although a single trait is sometimes also called a phenotype. For example, blond hair could be called a trait or a phenotype, as could obesity. A phenotype can be the result of many factors, including an individual's genotype, environment,

and lifestyle, and the interactions among these factors. The observed manifestation of a genotype. The phenotype may be expressed physically, biochemically, or physiologically.

Phylogeny (phylogenesis, phylogenetic, phylogenic): The evolutionary history of a particular taxonomic group, usually a species.

Plasmid: A structure composed of DNA that is separate from the cell's genome. In bacteria, plasmids confer a variety of traits and can be exchanged between individuals - even those of different species. Plasmids can be manipulated in the laboratory to deliver specific genetic sequences into a cell.

Polymerase chain reaction (PCR): The first practical system for in vitro amplification of DNA, and as such one of the most important recent developments in molecular biology. Two synthetic oligonucleotide primers, which are complementary to two regions of the target DNA (one for each strand) to be amplified, are added to the target DNA (that need not be pure), in the presence of excess deoxynucleotides and Taq polymerase, a heat-stable DNA polymerase. In a series (typically 30) of temperature cycles, the target DNA is repeatedly denatured (around 90°C), annealed to the primers (typically at 50-60°C) and a daughter strand extended from the primers (72°C). As the daughter strands themselves act as templates for subsequent cycles, DNA fragments matching both primers are amplified exponentially, rather than linearly. The original DNA need thus be neither pure nor abundant, and the PCR reaction has accordingly become widely used not only in research, but in clinical diagnostics and forensic science.

Polymorphism: in this context, the existence of inter-individual differences in DNA sequences coding for one specific gene. The effects of such differences may vary dramatically, ranging from no effect at all to the building of inactive proteins.

Primer: Short pre-existing polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase.

Probe: Single-stranded DNA or RNA molecules of specific base sequence, labelled either radioactively or immunologically, that are used to detect the complementary base sequence by hybridisation.

Promoter: a segment of DNA located at the 'front' end of a gene, which provides a site where the enzymes involved in the transcription process can bind on to a DNA molecule, and initiate transcription. Promoters are critically involved in the regulation of gene expression.

Proteome: total protein complement expressed by a cell, tissue or organism.

Proteomics: study of protein properties on a large scale to obtain a global, integrated view of cellular processes including expression levels, post translational modifications, interactions and location.

Randomly Amplified Polymorphic DNA (RAPD): Variation of the polymerase chain reaction used to identify differentially expressed genes. mRNA from two different tissue samples is reverse transcribed, then amplified using short, intentionally nonspecific primers. The array of bands obtained from a series of such amplifications is run on a high resolution gel and compared with analogous arrays from different samples. Any bands unique to single samples

are considered to be differentially expressed; they can be purified from the gel, and sequenced and used to clone the full-length cDNA.

Recombinant DNA: DNA molecules that have been created by combining DNA more than one source.

Regulatory Gene: a gene which controls the protein-synthesising activity of other genes.

Restriction enzymes (restriction endonucleases): a class of bacterial enzymes that cut DNA at specific sites.

Reverse Transcriptase: An enzyme used by retroviruses to form a complementary DNA sequence (cDNA) from an RNA template -usually the genome of the retrovirus. The enzyme then performs a complimentary template of the cDNA strand such that a double stranded DNA molecule is formed. This double stranded DNA molecule is then inserted into the chromosome of the host cell which has been infected by the retrovirus. Reverse transcriptase is one of the key components that HIV uses to mount its attack.

Ribosome: Small particles that are found in all cells both freely in the cytoplasm and more commonly found on endoplasmic reticulum. The ribosome particle consists of protein and ribosomal RNA, where the RNA is translated from the mRNA transcribed from nucleus DNA. This effectively is the point of creation of a protein, where each nucleotide in the polynucleotide is translated into part of the nucleotide sequence for the completed protein.

RNA (ribonucleic acid): a molecule similar to DNA, which helps in the process of decoding the genetic information carried by DNA.

RNAse (= ribonuclease, RNAase): Widely distributed type of enzyme that cleaves RNA. May act as endonucleases or exonucleases depending upon the type of enzyme.

RT-PCR (= reverse transcriptase polymerase chain reaction; reverse transcription PCR): PCR in which the starting template is RNA, implying the need for an initial reverse transcriptase step to make a DNA template. Some thermostable polymerases have appreciable reverse transcriptase activity; however, it is more common to perform an explicit reverse transcription, inactivate the reverse transcriptase or purify the product, and proceed to a separate conventional PCR.

Serum-responsiveness: cell proliferative reaction to the addition of serum to tissue culture medium after prior deprivation.

Sequencing: determining the order of nucleotides in a DNA or RNA molecule, or determining the order of amino acids in a protein.

Single Nucleotide Polymorphism (SNP): Inter-individual variations in the genetic code at the level of one nucleotide.

Southern Blotting: Transfer by absorption of DNA fragments separated in electrophoretic gels to membrane filters for detection of specific base sequences by radiolabeled complementary probes.

Species: Groups of populations (which are groups of individuals living together that are separated from other such groups) which can potentially interbreed or are actually interbreeding, that can successfully produce viable, fertile offspring (without the help of human technology).

Splicing: the removal of introns from the sequence of mRNA. When an mRNA molecule is synthesized from a DNA template, introns are transcribed (see transcription) along with exons. In the splicing process, this material is cut out and the exons are joined together to form a continuous coding sequence.

Slot blot: A dot blot in which samples are placed on a membrane through a series of rectangular slots in a template. This is slightly advantageous because hybridization artefacts are usually circular.

Structural genomics: The effort to determine the 3D structures of large numbers of proteins using both experimental techniques and computer simulation.

Subspecies: A group of organisms that is geographically isolated from and may display some morphological differences from other populations of a species, but is nevertheless able to interbreed with other such groups within the species where their ranges overlap.

Suppressor Gene: a gene which helps to reverse the effects of damage to an individual's genetic material, typically effects which might lead to uncontrolled cell growth (as would occur in cancer). A suppressor gene may, for example, code for a protein which checks genes for misspellings, and/or which triggers a cell's self-destruction if too many genetic mutations have accumulated.

Taq polymerase: A heat-stable DNA polymerase that is normally used in the polymerase chain reaction. It was isolated from *Thermus aquaticus*.

Toxicogenomics: a new scientific subdiscipline that combines the emerging technologies of genomics and bioinformatics to identify and characterize mechanisms of action of known and suspected toxicants. Currently, the premier toxicogenomic tools are the DNA microarray and the DNA chip, which are used for the simultaneous monitoring of expression levels of hundreds to thousands of genes.

Transcription: the process during which the information in a length of DNA is used to construct an mRNA molecule.

Transcriptomics: techniques available to identify mRNA from actively transcribed genes.

Transcriptome: mRNA from actively transcribed genes

Transcript Profiling: see transcriptomics

Transfer RNA (tRNA): RNA molecules which bond with amino acids and transfer them to ribosomes, where protein synthesis is completed.

Transformation: A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome.

Transgenic: An organism whose genome has been altered by the inclusion of foreign genetic material. This foreign genetic material may be derived from other individuals of the same species or from wholly different species. Genetic material may also be of an artificial nature. Foreign genetic information can be added to the organism during its early development and incorporated in cells of the entire organism. As an example, mice embryos have been given the gene for rat growth hormone allowing mice to grow into large adults. Genetic information can also be added later in development to selected portions of the organism. As an example, experimental genetic therapy to treat cystic fibrosis involves selective addition of genes responsible for lung function and is administered directly to the lung tissue of children and adults. Transgenic organisms have been produced that provide enhanced agricultural and pharmaceutical products. Insect resistant crops and cows that produce human hormones in their milk are just two examples.

Transgenic Organism: an organism whose genome has been altered by the incorporation of foreign, or exogenous DNA.

Translation: the process during which the information in mRNA molecules is used to construct proteins.

Vasopressin (antidiuretic hormone, ADH): A peptide hormone that increases blood pressure and the rate at which the kidneys absorb water, and is therefore used as an antidiuretic. It is secreted by the hypothalamus and stored in/released by the posterior pituitary gland.

Vector: [1] An organism which serves to transfer a disease causing organism (pathogen) from one organism to another. [2] a mechanism whereby foreign gene(s) are moved into an organism and inserted into that organism's genome. Retroviruses such as HIV serve as vectors by inserting genetic information (DNA) into the genome of human cells. Bacteria can serve as vectors in plant populations.

Vitellogenin: A protein, precursor of several yolk proteins, especially phosvitin and lipovitellin in the eggs of various vertebrates, synthesized in the liver cells after oestrogen stimulation. Also found in invertebrates.

Xenobiotic(s): substances not normally present in the reference organism

APPENDIX B REFERENCES

Anderson, H.R., Bland, M.J., Ponce de Leon, A., Strachan, D., Bower, J.S., 1991. Thorax 50, 1188-1193.

Arukwe, A., Celius, T., Walther, B.T., Goksoyr, A, 2000. Aquat. Toxicol. 49, 159-170.

Arukwe, A., Knudsen, F.R., Goksoyr, A., 1997. Environ. Health Perspect. 105, 418-422.

Arukwe, A., Kullman, S.W., Hinton, D.E., 2001a. Comparative Biochemistry and Physiology C-Toxicology & Pharmacology 129, 1-10.

Arukwe, A., Yadetie, F., Male, R., Goksoyr, A., 2001b. Environmental Toxicology and Pharmacology 10 (1-2), 5-15.

Atienzar, F.A., Conradi, M., Evenden, A.J., Jha, A.N., Depledge, M.H., 1999. Environmental Toxicology and Chemistry 18 (10), 2275-2282.

Begovac, P.C., Wallace, R.A., 1989. J. Exp. Zool. 251,56-73.

Bortone, S.A., Davis, W.P., Bundrick, C.M., 1989. Bull Environ Contam Toxicol 43, 370-377.

Celius, T., Matthews, J.B., Giesy, J.P., Zacharewski, T.R., 2000. Journal of Steroid Biochemistry and Molecular Biology 75 (2-3), 109-119.

Celius, T., Walther, B.T., 1998a. J. Endocrinol. 158, 259-266.

Celius, T., Walther, B.T., 1998b. J. Exp. Zool. 281, 346-353.

Chen, C.Y., Sillett, K.B., Folt, C.L., Whittemore, S.L., Barchowsky, A., 1999. Hydrobiologia 401, 229-238.

Chen, T.T., 1983. Can. J. Biochem. Cell Biol. 61, 802-810.

Christiansen, L.B., Povlsen, A., Pedersen, S.N., Korsgaard, B., Bjerregaard, P., May 21-25 2000. In Third SETAC World Congress, Brighton, United Kingdom, pp. 136.

Christiansen, T., Korsgaard, B. and Jespersen, Å., 1998. Mar. Environ. Res. 46:141-144.

Conte, C., Mutti, I., Puglisi, P., Ferrarini, A., Regina, G., Maestri, E., Marmiroli, N., 1998. Chemosphere 37 (14-15), 2739-2749.

Cotteli, F., 1998. In: Proceedings of the EMBO workshop on reproduction and development, Bergen, Norway. P42.

Crane, M., Byrns, G., 2002. R&D Technical Report P300, Environment Agency, Bristol.

Culotta, V.C., Lin, S.J., Schmidt, P., Klomp, L.W.J., Casareno, R.L.B., Gitlin, J., 1999. Advances in Experimental Medicine and Biology 448, 247-254.

Denslow, N.D., Bowman, C.J., Ferguson, R.J., Lee, H.S., Hemmer, M.J., Folmar, L.C., 2001a. General and Comparative Endocrinology 121, 250-260.

Denslow, N.D., Lee, H.S., Bowman, C.J., Hemmer, M.J., Folmar, L.C., 2001b. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 129 (2-3), 277-282.

Department of Health (1995), Committee on the Medical Effects of Air Pollutants: Particles, 36-37.

Desbrow, C., Routledge, E., Brighty, G.C., Sumpter, J.P., Waldock, M., 1998. Environ Sci Technol 32,1549-58.

di Toppi, L.S., Gabbrielli, R., 1999. Environmental and Experimental Botany 41, 105-130.

Evans, C.W., Wilson, D.A., Mills, G.N., 2001. Biomarkers 6, 7-14.

Flammarion, P., Brion, F., Babut, M., Garric, J., Migeon, B., Noury, P., Thybaud, E., Tyler, C.R., Palazzi, X., 2000. Ecotoxicology 9,127-135.

Flouriot, G., Pakdel, F., Ducouret, B., Valotaire, Y.J., 1995. Mol. Endocrinol. 15, 143-51.

Floyd, R., Abebe, E., Papert, A., Blaxter, M., 2002. Molecular Ecology 11, 839-850.

Folmar, L.C., Hemmer, M., Hemmer, R., Bowman, C., Kroll, K., Denslow, N.D., 2000. Aquat. Toxicol. 49, 77-88.

Fujita, S., Chiba, I., Ishizuka, M., Hoshi, H., Iwata, H., Sakakibara, A., Tanabe, S., Kazusaka, A., Masuda, M., Masuda, Y., Nakagawa, H., 2001. Biomarkers 6, 19-25.

Gagne, F., Bliase, C., 1998. Aquatic Toxicol. 44: 83-91

Gallagher, E.P., Sheehy, K.M., 2000. Marine Environmental Research 50 (1-5), 399-403.

Halm, S., Rand-Weaver, M., Sumpter, J.P., Tyler, C.R., 2002. Aquatic Toxicol. 60: 285-299.

Hamazaki, T.S., Nagahama, Y., Iuchi, I., Yamagami, K., 1989. Dev.Biol. 133, 101-110.

Harries, J., Jambakhsh, A., Jobling, S., Matthiessen, P., Sumpter, J.P., Tyler, C.R., 1999. Environ. Toxicol. Chem. 18 (5), 932-937.

Harris, C., Santos, E.M., Pottinger, T.G., Tyler, C.R., Sumpter, J.P., 2001. Environ Sci Technol.

Hashimoto, S., Bessho, H., Hara, A., Nakamura, M., Iguchi, T., Fujita, K., 2000. Mar Environ Res 49,37-53.

Hecker, M., Tyler, C.R., Maddix, S., Karbe, L., 2002. Environ Sci Technol. In press.

Hemmer, M.J., Bowman, C.J., Hemmer, B.L., Friedman, S.D., Marcovich, D., Kroll, K.J., Denslow, N.D., 2002. Aquat. Toxicol. 58. In press.

Hemmer, M.J., Hemmer, B.L., Bowman, C.J., Kroll, K.J., Folmar, L.C., Marcovich, D., Denslow, N.D., 2001. Environ. Toxicol. Chem. 20, 336-343.

Hodgson, P.A., George, S.G., 1998. Marine Environmental Research 46 (1-5), 465-468.

Hopkin, S. P. Ecophysiology of Metals in Terrestrial Invertebrates; Elsevier Applied Science: London, UK, 1989.

Hopkin, S.P., 1993. In: Calow, P. (Ed.), Handbook of Ecotoxicology. Blackwell Scientific Publications, Oxford, pp. 397-427.

Huet M-C, 2000. Ecotoxicology 9: 77-84.

Jobling S., Nolan M., Tyler C.R., Brighty G., Sumpter J.P., 1998. Environmental Science & Technology 32(17):2498-506.

Jobling S., Tyler, C.R., 2003. In: Proceedings of the SCOPE/IUPAC International Meeting on Endocrine Active Substances, Japan. In press.

Jobling, S., Beresford, N., Nolan, M., Rodgers-Gray, T., Brighty, G.C., Sumpter, J.P., Tyler, C.R., 2002a. Biol Reprod 66(2),272-81.

Jobling, S., Coey, S., Whitmore, J., Beresford, N., Nolan, M., Brighty, G., Sumpter, J.P., Tyler, C.R., 2001. Biology of Reproduction. 66(2):272-281.

Jobling, S., Coey, S., Whitmore, J., Kime, D.E., vanLook, K.J.W., McAllister, B.G., Beresford, N., Henshaw, A.C., Brighty, G., Tyler, C.R., Sumpter, J.P., 2002b. Biol Reprod. In press

Jobling, S., Nolan, M., Tyler, C.R., Brighty, G., Sumpter, J.P., 1998. Environ Sci Technol 32, 2498-2506.

Jobling, S., Sheahan, D., Osborne, J.A., Matthiessen, P. and Sumpter, J.P., 1996. Environ. Toxicol. Chem. 15:194-202.

Kammenga, J., 2000. SETAC Europe 11th Annual Meeting, Madrid, pp. 47.

Kille, P., Sturzenbaum, S.R., Galay, M., Winters, C., Morgan, A.J., 1999. Pedobiologia 43, 602-607.

Kim S.K., Lund J., Kiraly M., Duke K., Jiang M., Stuart J.M., Eizinger A., Wylie B.N., Davidson G.S., 2001. Science, 293: 2087-2092.

Kiyono, M., Uno, Y., Omura, T., Pan-Hou, H., 2000. Journal of Health Science 46, 142-145.

Kochanowski, B., Reischl, U. (Eds.), 1999. Quantitative PCR Protocols. Humana Press, Totowa, New Jersey.

Kumar, T.R., Prasad, M.N.V., 1999. Journal of Plant Physiology 155 (4-5), 652-655.

Larsson, D.G.J., Adolfsson Erici, M., Parkkonen, J., Pettersson, M., Berg, A.H., Olsson, P.E., Forlin, L., 1999. Aquat Toxicol 45,91-97.

Li, Y. M.; Chaney, R. L.; Siebielec, G.; Kerschner, B. A. J. Environ. Qual. 2000, 29, 1440-1447.

Lye, C.M., Frid, C.L.J., Gill, M.E., McCormick, D., 1997. Mar Pollut Bull 34, 34-41.

Maack, G., 2002. PhD thesis. Leipzig, Germany.

Maywald, F., Weigel, H.J., 1997. Landbauforschung Volkenrode 47, 103-126.

Meharg, A. A.; Osborn, D.; Pain, D. J.; Sanchez, A.; Naveso, M. A. Environ. Pollut. 1999, 105, 387-390.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Cold Spring Harbor Symposia on Quantitative Biology 51, 263-273.

Munkittrick, K.R., Vanderkraak, G.J., 1994. Pulp & Paper-Canada, 95(5), 57-59.

Murphy, S.A., Bérubé, K.A., Pooley, F.D., Richards, R.J., 1998. Life Sci. 62, 1789-1799.

Muyzer, G., Dewaal, E.C., Uitterlinden, A.G., 1993. Applied and Environmental Microbiology 59, 695-700.

Nogawa, K.; Teruhiko, K. In Toxicology of Metals; Magos, L., Suzuki, T., Eds.; CRC Press Inc.: Boca Raton, 1996; pp 353-369.

Morgan, A.J., Stürzenbaum, Winters, C., Grime, G.W., Abd.Aziz, N.A., Kille, P, 2003. Ecotoxicology. In press.

Morgan, A.J., 2003. Metals, Earthworms and Vermicompost: A Review Based on Current Regulations for Land Disposal of Sewage Sludge. In press

Nicholson J.K., Connelly J., Lindon J.C., Holmes E., 2002. Nature Reviews Drug Discovery 1(2), 153-161.

Oppen-Bernsten, D.O., Gram-Jensen, E., Walther,B.T., 1992. J. Endocrinol. 135, 293-302.

Organization for Economic Cooperation and Development. Final report of the fish expert consultation meeting, London, 28-29 October 1998. Environmental Health and Safety Division, Paris, France, Report no. 9906. (1999)

Osier, M., Oberdörster, G., 1997. Fundam. Appl. Toxicol. 40, 220-227

Pakdel, F., Feon, S., LeGac, F., LeMenn, F., Valotaire, Y., 1991. Mol. Cell. Endocrinol. 75, 205-212.

Pathiratne, A., George, S.G., 1998. Aquatic Toxicology 43, 261-271.

Peakall, D., 1992. Animal Biomarkers and Pollution Indicators. Chapman and Hall, London.

Pennington, S.E., Dunn, M.J. (Eds.), 2001. Proteomics: From Protein Sequence to Function. BIOS Scientific Publishers Limited, Oxford, UK.

Piferrer, F., Donaldson, E.M., 1992. Aquaculture 106:183-193.

Purdom, C.E., Hardiman, P.A., Bye, V., Eno, N.C., Tyler, C.R., Sumpter, J.P., 1994. Chemistry and Ecology. 8:275-285.

Richards, J.R., Cryer, J., Kille, P., 2000. Molecular genetic response of Daphnia to environmental contaminants, SETAC Third World Congress, Brighton, UK.

Richards, R.J., 1997. Biologist 44, 249-251

Richards, R.J., Bérubé, K.A., Masek, L., Symons, D., Murphy, S.A., 1999. In: Maynard, R.L., Howard, C.V. (Eds.), Particulate Matter: Properties and Effects upon Health. BIOS Scientific Publishers Ltd, New York, pp.97-113.

Rigina, O.; Baklanov, A.; Hagner, O.; Olsson, H. Sci. Total Environ. 1999, 229, 147-163.

Risso-de Faverney, C., Lafaurie, M., Girard, J.P., Rahmani, R., 2000. Environmental Toxicology and Chemistry 19, 2239-2248.

Rodgers-Gray, T.P., Jobling, S., Kelly, C., Morris, S., Brighty, G., Waldock, M., Sumpter, J.P., Tyler C.R., 2001. Environ Sci and Technol. 35:(3) 462-470.

Rodgers-Gray, T.P., Jobling, S., Morris, S., Kelly, C., Kirby, S., Janbakhsh, A., Harries, J.E., Waldock, M.J., Sumpter, J.P., Tyler, C.R., 2000. Environ Sci Technol 34:1521-1528.

Sarkar, B. (Ed.), 1999. Metals and Genetics. Kluwer Academic, New York.

Savva, D., 2000. Polycyclic Aromatic Compounds 20 (1-4), 291-303.

Schlezinger, J.J., Stegeman, J.J., 2000. Aquatic Toxicology 50, 375-386.

Schmid,T., Gonzalez-Valero, J., Rufli, H., Dietrich, D.R., 2002. Toxicol Letts 131,65-74.

Scholz, S., Gutzeit, H.O., 2000. Aquat. Toxicol. 50, 363-373.

Schwartz, J., 1994. Environ. Res. 64, 26-35.

Sethi, S.; Grover, S.; Khodaskar, M. B. Annals Tropical Paediatrics 1993, 13, 3-6.

Somerfield, P. J.; Gee, J. M.; Warwick, R. M. Marine Pollution Bulletin, 1994, 28, 363-369.

Spurgeon, D.J., Svendsen, C., Hankard, P.K., Weeks, J.M., Kille, P., Fishwick, S.K., 2002. Review of Sublethal Ecotoxicological Tests for Measuring Harm in Terrestrial Ecosystems. R&D Technical Report P5-063/TR1, Environment Agency, Bristol.

Spurgeon., D. S., Stürzenbaum., S. R., Svendsen, C., Hankard., PK., Morgan., A. J., Weeks, J. M. and Kille, P. 2003 ES&T submitted

Stackebrant E. & Goodfellow, M. (1991). Nucelic acid techniques in bacterial systematics. John Wiley & Sons, Chichester, England.

Sturmbauer, C., Opadiya, G.B., Niederstatter, H., Riedmann, A., Dallinger, R., 1999. Molecular Biology and Evolution 16, 967-974.

Stürzenbaum, S.R., Cater, S., Morgan, A.J., Kille, P., 2001. Biometals 14, 85-94.

Stürzenbaum, S.R., Kille, P., 2001. Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology 130, 281-289.

Stürzenbaum, S.R., Kille, P., Morgan, A.J., 1998. FEBS Letters 431, 437-442.

Stürzenbaum, S.R., Winters, C., Galay, M., Morgan, A.J., Kille, P., 2001. J. Biol. Chem. 276, 34013-34018.

Tao, H., Bausch, C., Richmond, C., Blattner, F.R. Conway T., 1999. J. Bacteriol. 181, 6425-6440

Theodorakis, C.W., Shugart, L.R., 1998. Ecotoxicology 7, 227-236.

Thomas-Jones, R.E.,Thorpe K.L, Harrison, N., Thomas, G., Morris C.A., Hutchinson, T.H., Woodhead, S., Tyler, C.R., 2003. Environ. Toxicol. Chem. Submitted.

Thorpe, K., Brighty, G., Cummings, R., Hutchinson, T.H., Scholze, M., Sumpter, J.P., Tyler, C.R., 2002. Environ Science and Technol. Submitted.

Thorpe, K., Hutchinson, T.H., Hetheridge, M., Sumpter, J.P., Tyler, C.R., 2000. Environ Toxico Chem. 19: 2812-2820.

Thorpe, K., Hutchinson, T.H., Hetheridge, M., Sumpter, J.P., Tyler, C.R., 2000. Environ Toxicol and Chem. 19: 2812-2820.

Thorpe, K.L., Hetheridge, M., Hutchinson, T.H., Scholze, M., Sumpter, J.P., Tyler, C.R., 2001. Environ Sci Technol 35,2476-2481.

Toft, G., Baatrup, E., 2001. Ecotox. Environ. Saf. 48:76-84.

Tully, D.B., Collins, B.J., Overstreet, J.D., Smith, C.S., Dinse, G.E., Mumtaz, M.M., Chapin, R.E., 2000. Toxicology and Applied Pharmacology 168, 79-90.

Tyler, C.R., Jobling S., Sumpter, J.P., 1998. Critical Reviews in Toxicology 28(4):319-61.

Tyler, C.R., Routledge, E.J., 1998. Pure & Appl Chem 70(9),1795-1804.

Tyler, C.R., Sumpter, J.P., Witthames, P.R., 1990. Biol. Reprod. 43, 202-209.

US Environment Protection Agency. Endocrine Disruptor Screening Programme. Proposed Strategy Statement of Policy. Fed Reg pp71542-71568 (1998).

Valencia R., Gerpe M., Trimmer J., Buckman T., Mason A.Z., Olsson P.E., 1998. Marine Environmental Research 46 (1-5): 167-171.

van Aerle, R., Nolan, M., Jobling, S., Christiansen, L.B., Sumpter, J.P., Tyler, C.R., 2001. Environ Toxicol Chem. In press.

van Aerle, R., Pounds, N., Maddix, S., Hutchinson, T.H., Tyler, C.R., 2002. Ecotoxicology. Accepted. In press

Vos, J.G., Dybing, E., Greim, H.A., Ladefoged, O., Lambŕe C., Tarazona, J.V., Brandt, I., Vetthaak, A.D., 2000. Crit Rev Toxicol 30,71-133.

Wagner M, Amann R, Lemmer H, Schleifer K-H. (1993). Probing activated sludge with oligonucleotides specific for proteobacteria: inadequacy of culture dependent methods for describing microbial community structure. Applied Environmental Microbiology 59: 1520-1525.

Welsh Office, 1988. Survey of Contaminated land in Wales. Environmental Advisory Unit, the Welsh Office (ISBN 086348 8285).

Woese, C R. (1987). Bacterial Evolution. Microbiological reviews 51: 221-271.

APPENDIX C NERC FUNDED PROGRAMMES

AWARDS MADE IN ROUND ONE AND ROUND TWO OF THE PROGRAMME

C.1 Molecular evolution of genes controlling vernalization requirement and response

Dr C Dean.

The availability of the Arabidopsis thaliana genome sequence together with an emerging understanding of the genetic pathways that regulate flowering time has opened up the potential to analyse the molecular evolution of the genes underlying flowering time variation and to investigate its selective importance. We will focus on variation in vernalization requirement and response and aim to characterize molecular variation in known and new loci, ascertaining its role in phenotypic variation. We will also aim to understand the ecological significance of the phenotypic variation caused by defined molecular changes and examine the selective forces on the alleles through examination of the pattern of sequence polymorphism at and surrounding the loci.

C.2 Data mining the Fugu genome sequence to characterise genes regulating muscle growth in a marine fish

Prof I A Johnston.

The phenotypic plasticity of muscle growth will be investigated in relation to temperature in the marine fish Fugu rubripes at the embryo and juvenile stages of the life cycle using an observational and hypothesis testing approach. Bioinformatics techniques will be used to mine sequence information from the genome sequence to identify orthologues of gene loci involved in muscle growth in mammals. Computational analysis will be applied to interrogate gene promoter regions to identify muscle-related cis-acting regulatory sites on a genome-wide basis. As a complimentary approach, subtracted cDNA libraries prepared from embryo/larval and adult stages will be used to identify candidate genes involved in muscle fibre recruitment. The expression patterns of novel and known genes will be investigated using competitive PCR and in situ hybridisation.

C.3 Genetic control of plastic and evolutionary responses to the environment in Drosophila melanogaster

Prof L Partridge.

Phenotypic plasticity, short-term evolutionary responses to selection and geographical clines in the responses of Drosophila melanogaster to changes in temperature and parasitoid attack will be studied using whole-genome and customised DNA microarrays. The data will allow the genes and mechanisms controlling the responses to these two important environmental challenges to be identified and understood in a way hitherto impossible.

C.4 The (Meta)Genomic Diversity of the Nitrogen Cycle - A New Approach to an Old Experiment

Prof AWB Johnston.

We will make Metagenomic Libraries, using fragments, ca 30 kb in size, cloned in the wide host-range cosmid vector, pLAFR1. The DNA will be obtained directly from bacteria in soils at the unique Rothamsted Broadbalk Experiment which differ in their N fertiliser regimes over the past 170 years. Libraries will be screened for functional genes for several steps in the N-Cycle by transferring them into appropriate mutants of various bacterial taxa. The resultant genes and flanking sub-genomes will be analysed. Thus, we will identify novel, processes, genes and genomes involved in this important biogeochemical cycle, even in bacteria that cannot be cultured. We will also determine the effects of long-term environmental differences on biodiversity of soil bacteria, at the levels of the genes, the processes, the genomes and the taxa that are responsible for driving the N-cycle.

C.5 Defining the bacterial accessory gene pool

Prof JPW Young.

The accessory genome of bacteria includes genes that may confer important adaptations in certain environments, may have a distinctive base composition. are often carried on mobile genetic elements, and may be more readily transferred between bacteria than the basic genome. We will exploit the published complete genome sequence of a Mesorhizobium strain and develop a high-throughput methodology based on DNA hybridisation and melting to describe the differences between accessory and basic genomes in this genus and related bacteria. We will monitor divergence of each of about 500 genes in about 100 bacterial isolates, and develop a general view of bacterial population genomics based on these data and published genome data.

C.6 Exploiting genomic information: strategies for cross-species identification in proteomics

Prof RJ Beynon.

Although proteins have the advantage of being direct reporters of the outcome of gene expression, proteomics is conducted most simply when the target genome is known, a situation that rarely pertains in NERC-directed research. This programme is an exploratory study of proteomics applied to problems such as chemical communication between individuals, infection and population dynamics and sperm competition. The first part of the programme will be a critical assessment of the utility of different proteomics methods in the absence of genome data, and will explore a novel strategy for cross-species matching. In the second part of the programme, selected samples, derived from ongoing NERC programmes, will be analysed to assess the utility of the different matching approaches and to provide preliminary data on the role that proteomics might play in studying such problems.

C.7 Transcriptome and functional analysis of marine rhythmicity

Prof PJW Olive.

Much of behaviour and physiology is determined by responses to environmental variables that cycle with regular periodicities, such as the light-dark cycle, and almost all organisms have evolved biological oscillators that can anticipate these changes. Spectacular advances have been made in understanding the molecular basis of the circadian clock in various model terrestrial organisms. This is in stark contrast to marine organisms that express both circadian and lunar cycles. To address this imbalance, we shall study the molecular basis of behavioural cycles that are observed in the commercially important marine lophotrochozoan, Nereis virens. In addition we will probe the molecular switches that underlie changes in the reproductive behaviour and physiology of this organism, and which is under photoperiodic control. This shall be done by identifying the Nereis homologues of the canonical clock genes that are found in insects and mammals, and studying their expression patterns, both temporally and spatially. In addition, a transcriptomics approach will be taken in order to identify novel genes, whose expression patterns correlate with behavioural cycling, or with changes in the reproductive programme. As Nereis populations show considerable variation in these behavioural and physiological phenotypes that correlate with regional environmental differences, differential expression of these novel genes in these populations, may reflect their role in the adaptation of each population to its particular geographic location.

C.8 Application of micro-array technology to reveal the complexity and genetic basis of the environmental adaptive osmoregulatory physiology of euryhaline teleosts

Prof R Balment.

The ability of some fish to survive in both freshwater and seawater represents one of the most extreme examples of adaptive physiology in the animal kingdom. This project will identify novel genes which are differentially expressed in response to hypo- and hyper-osmotic challenge in teleost fish. This will be achieved by the screening of micro arrays constructed for two native euryhaline species, the eel and flounder. In a second phase the project will examine in detail a selected number of novel differentially expressed genes using molecular, biochemical and physiological techniques to define the function of these systems in relation to osmotic changes in the environment. The results will provide a more complete picture of the mechanisms and pathways involved in the control of osmoregulation in teleost fish to be constructed.

C.9 Genomic analysis of the relationship between repetitive DNA and phenotypic evolution

Prof TR Meagher.

One revelation of genomics is the evolutionary importance of DNA content composition as well as sequence. Our previous work on Silene latifolia showed that variation in nuclear DNA content, attributable to non-coding DNA, was negatively correlated with flower size. Other recent investigations have explored the composition of non-coding DNA elements which have been found to fall into a range of different types (e.g., microsatellites, retrotransposons, etc.).

We plan to identify, by means of in situ hybridisation analysis of the S. latifolia genome, the relative contributions of different types of repetitive elements to phenotype variation. This study will provide new insight into the role of non-coding DNA in phenotypic evolution.

C.10 Genomic and proteomic analyses of adaptive evolution in experimental populations of Pseudonomas fluorescens

Dr P Rainey.

Understanding, at a mechanistic level, the moment-by-moment workings of adaptive evolution is essential for biological science. However, the challenges posed are immense and progress requires simple experimental populations where evolutionary processes can be observed in real time. Using populations of the bacterium Pseudomonas fluorescens, we propose a multifaceted programme of research that combines state of the art proteomic analyses, whole-genome mutation screening protocols and biochemical studies of signal transduction pathways. Together, these technologies and their innovative application will generate a comprehensive account of a single moment of adaptive evolution - from primary DNA sequence change, through to form and fitness.

C.11 Environmental fitness of Burkholderia cepacia: feasibility of testing with the genomic approach of signature tagged transposon mutagenesis

Dr E Mahenthiralingham.

The Burkholderia cepacia complex is a closely related group of free-living bacteria which are ubiquitous in the natural environment. They possess useful bioremediation and biological control properties, but, in contrast, are capable of opportunistic human infection. We will attempt to determine the genetic basis for the environmental fitness of B. cepacia. A modern genetic technology, signature tagged mutagenesis (STM), un-tested in an ecological setting, will be assessed in simple laboratory models of B. cepacia rhizosphere colonization and pollutant degradation. STM is a genomic approach enabling the identification of global gene expression and protein function required for microbial adaptation to diverse habitats. This fundamental research will greatly enhance our understanding of ecological impact and health risks associated with the release of microbial biotechnological agents.

C.12 Changes to genome and transcriptome during hybrid speciation

Dr SJ Hiscock.

This project will examine changes in genome structure and transcriptome during abrupt hybrid speciation in the genus Senecio. Research will consist of two priority genomic analyses, 1) the homoploid hybrid origin of S. squalidus, focussing on genome reorganization and gene expression, and 2) the allopolyploid origin of S. cambrensis, focussing on genome reorganization, gene expression and phenotypic evolution, we will compare these processes in both natural (wild) hybrids and resynthesized hybrids. Each genomic analysis will consist of parallel AFLP/SSR mapping (at St. Andrews) and micro array analysis (at Bristol) of parental and hybrid genomes/transcriptomes. ESTs showing altered patterns of expression will be

characterized and positioned on the saturated maps to investigate correlations between changes in genome structure and changes in gene expression.

C.13 Environmental genomics of calcicole-calcifuge physiology

Dr MR McAinsh.

Calcareous grasslands are among the most species-rich plant communities in Europe. A clear understanding of the molecular mechanisms that enable calcicole species to thrive on calcareous soils will contribute to the development of strategies to maintain the biodiversity of these vulnerable ecosystems. We will test the hypothesis that adaptation to a calcareous environment reflects altered patterns of gene expression. We will identify genes that are differentially expressed between calcicolous and non-calcicolous ecotypes of A. thaliana and determine their role in the calcicole phenotype. We will examine whether these calcicole adaptation genes show similar patterns of expression in wild A. thaliana and related Arabs species from different calcareous habitats and determine the impact of environmental stress on their expression and on the calcicole phenotype.

C.14 Assessing contributions to fitness of individual genes via genome-wide competition analysis

Prof SG Oliver & Prof DB Kell.

In unicellular microbes, fitness = growth rate, and very small changes in growth rate, and thus fitness, may be discriminated using competition analysis in chemostat culture. Molecularly bar-coded strains of bakers yeast containing all possible (Ca 6000) single-gene deletions will be subjected to competition analysis (250 strains at a time) in chemostats limited in 6 different ways. The growth rates and gene expression profiles will be analysed, respectively, by transcriptome arrays (a) newly produced for the bar-codes and (b) those existing under COGEME for every mRNA/cDNA. Subsets will be separated by flow cytometric cell sorting of strains stained for viability and levels of membrane energisation and similarly analysed. Data will be stored in the Manchester Genome Information Management System. Genetic programming methods will be used to provide optimal rules describing the relationship between gene expression levels and fitness in each environment. Other environments chosen will include those containing toxic metals and xenobiotics, oxidative stress, osmotic stress conditions, and the presence of weak acids (such as those used as preservatives for fruit juices). Finally, strains will be studied in their natural environments, in juices, rotting fruit and in oak tree exudates. This unique genetic resource will provide an unparalleled opportunity to assess the global contribution to fitness of every gene in a single organism.

C.15 A genomic approach to investigate infection dynamics of marine phytoplankton viruses

Dr Willie Wilson.

Viral lysis of marine phytoplankton is recognised as playing a key role in global biogeochemical cycles. Advances in genomic technology have presented us with a new toolbox of techniques that we propose to use to investigate the unique life histories of phytoplankton viruses and help resolve why marine micro-algal viruses have the largest virus

genomes known. We will develop an Emiliania huxleyi virus microarray and use transcriptomics to determine function of an estimated 400 - 500 genes in its 400 kbp genome. Fieldwork will be conducted to investigate expression of virus latency genes in E. huxleyi blooms. A bacterial artificial chromosome (BAC) virus genomic library will be generated from a natural virus community to determine life histories of unculturable viruses through sequence analysis.

C.16 Exploiting terrestrial sentinel and model species to integrate tiers of biological response to pollution

Dr Pete Kille.

To use 'omic' technologies (transcriptomics, proteonics, mechanistics and demographics) to catalogue the response cascades in model and sentinel soil organisms exposed to a range of pollutants of concern. Central to achieving this goal will be the exploitation of the vast bioinformatic resources available for a model organism (the nematode C. elegans) to identify and characterise pollutant responses shared with a sentinel invertebrate (the earthworm L. rubellus). This unique data set will identify key pollutant response pathways consistent across taxa and will allow the genotypic, phenotypic and ecological consequence of soil pollution to be linked.

C.17 Genetic control of avian plumage traits: towards the mechanisms underlying sexual selection

Professor Terry Burke.

This project will identify genes and pathways controlling feather pigmentation, patterning, growth and distribution, and zone-specific expression. We will: (1) Identify genes expressed in avian skin, (2) Prepare a microarray comprising these genes, (3) Compare expression profiles between different feather zones within birds and between different genotypes (including sex), (4) Exploit time-series analysis to understand sequential gene regulation, (5) Confirm the regulation of identified candidate genes, (6) Test the utility of the microarrays in a range of bird species including phasianids and passerines, and (7) Use comparative analyses to test the role of sexual selection in promoting the evolution of plumage genes.

C.18 Molecular bases of zinc tolerance and accumulation by Arabidopsis halleri

Professor MR Macnair.

Transcriptional profiling of the tolerant, accumulating species Arabidopsis halleri, the non-accumulating and non-tolerant species A. petraea, A. thaliana and A. neglecta, and F3 families of the cross between A. petraea and A. halleri will be conducted under various zinc treatments. Genes whose expression is correlated with either tolerance or accumulation will be used as RFLP markers in a QTL analysis of a segregating F2 family of the A. petraea x A. halleri cross. QTL analysis will also be used to determine whether there are any genes which are not detected by transcriptional profiling. Alleles of adaptive genes identified by this strategy will be sequenced from A. halleri and its sister species.

C.19 Comparative genome architecture of the hominids - biological, evolutionary and ecological significance

Dr PH Dear.

The aim of this project is a detailed comparison of the genomes of all of the hominid species - chimpanzee, bonobo, gorilla, orangutan and man. The project has been designed to reveal the changes in genome architecture and organisation that are believed to play a key role in speciation. We will use a uniquely appropriate technique to construct, for the first time, genome maps of all the species at progressively higher resolution, revealing both the general nature and specific instances of chromosomal rearrangements. We will also survey the genomes for duplications, and correlate these with differential gene expression. Our findings will both complement and support genome sequencing projects, and are of direct relevance to basic biology, to biodiversity and conservation, biomedicine, and to the mechanisms of evolution and speciation.

C.20 Identification of the mechanisms of disruption of sexual differentiation in fish exposed to EDCs and their mixtures: a transcriptomic approach

Professor CR Tyler

Endocrine disrupting chemicals (EDCs) in STWs effluents cause intersex in wild indigenous fish, which results in reduced reproductive output and potentially has population-level consequences. A major challenge now is to unravel how EDCs impact on the genome of organisms to disrupt sexual development to inform on pathways of action of EDCs and direct the development of systematic screens for these chemicals. We will determine the cascades governing gonad sexual differentiation in the roach (Rutilus rutilus) by characterising the transcriptomic and whole organism responses in fish exposed to known EDCs and their mixtures. Complimentary studies will be undertaken in the zebrafish (ZF, Danio rerio) to assess the applicability of the ZF for informing potential impacts of EDCs on sexual development in wild, indigenous, cyprinid fish.

C.21 QTL mapping in natural populations using linkage and linkage disequilibrium

Dr JM Pemberton

We will investigate the genetic architecture underlying quantitative variation an selected morphometric and life history traits in a free-living population, the Soay sheep on St. Kilda. (1) We will conduct a full-genome scan for QTL using microsatellite markers and linkage mapping within a series of large, connected, sibships. (2) We will generate SNP markers in targeted regions around already-known candidate QTL, to investigate the relationship between linkage disequilibrium (LD) and genetic distance in order to determine the feasibility of LD mapping in the study population. (3) If LD mapping is feasible, we will fine-map other candidate QTL emerging from our own linkage mapping and from studies of domestic sheep.

C.22 The genetic control of phenotypic plasticity in the life-cycle of the free-living nematode Caenorhabditis elegans

Dr ME Viney.

In the life-cycle of C. elegans there is a developmental choice (dauer or non-dauer development), modulated by the worms' environment. This is an example of a phenotypically plastic trait, i.e. an environmentally induced change in phenotype. Isolates of C. elegans vary in their phenotypic plasticity of dauer development. We propose to identify the loci controlling this phenotypic plasticity by a combined analysis of gene expression (using whole genome microarrays) and mapping the quantitative trait loci. We will also investigate fitness consequences of variation in this trait. This work will use the extensive information, methods and genomic resources available with C. elegans to understand the genetic basis and phenotypic consequences of a plastic trait, which is central to understanding GENOTYPE x ENVIRONMENT interactions and how selection acts on them.

C.23 Population structure of a segmented RNA animal virus: immune mediated clonality or free reassortment

Nunn Dr MA.

The aim of this proposal is to understand the mechanisms that maintain pathogen diversity using a genetically and phenotypically polymorphic, globally distributed seabird obligate tick-borne orbivirus that has a segmented double-stranded RNA genome. The organism is particularly well suited to addressing issues of geographical and temporal scale and interaction between virus and host immunity. Permitting a rigorous empirical test of the widely debated theory of immune mediated clonality (Gupta et al., 1994a and 1994b). The molecular evolutionary history of each of the pathogens 10 genes will be investigated and direct estimates of sequence rate parameters obtained. As an additional benefit, the study design will enable an assessment of the impact of virus infection on host mortality and direct estimates of rates and routes of virus transmission in the natural host.

C.24 Global transcriptional effects of UV-B and insect plant stressors on interactions between Arabidopsis and rhizosphere bacteria

Dr RF Preziosi.

We will use a simple model ecosystem to build a global view at the transcriptome level of how interactions between plants (Arabidopsis) and rhizosphere bacteria (Pseudomonas) are altered by biotic stressors (aphids and caterpillars) and by abiotic stressors (UV-B). Plant and bacterial gene expression in stressed and unstressed ecosystems will be examined by whole-genome microarray analysis, and correlated with phenotypic and fitness changes in all three components of the insect-plant-rhizobacteria ecosystem. We anticipate the discovery of potential mechanistic pathways by which environment and global gene expression interact to affect fitness, and the identification of novel genes involved in plant-rhizosphere interactions. The project will also produce valuable new tools for the statistical treatment of environmental genomics data.

C.25 RpoN-dependent adaptation in Pseudomonas fluorescens

Dr GM Preston.

The alternative sigma factor standard deviation 54 is predicted to regulate a wide range of traits in Pseudomonas fluorescens SBW25. The full repertoire of genes regulated by standard deviation 54 is unknown. In order to understand the role of standard deviation 54 in P. fluorescens SBW25 it is necessary to identify the components and targets of the standard deviation 54 regulatory network; the effect of environmental factors on standard deviation 54 dependent transcription; and the potential for cross talk between standard deviation 54-activating proteins. Standard deviation 54-activating proteins and standard deviation 54-dependent promoters can be predicted from genome sequence data. These predictions will be used in conjunction with microarray technology and site-directed mutagenesis to generate, test and refine a model of the standard deviation 54 regulatory network. The explanatory power of this model will be tested directly by analysing gene expression and the performance of mutants in the natural environment.

C.26 Linkage disequilibrium and sequence polymorphism in S-locus haplotypes of the self-incompatible plant Arabidopsis lyrata

Professor D Charlesworth.

I propose an in depth study of a particularly interesting genome region, the region containing the Arabidopsis lyrata pistil and pollen incompatibility genes (SRK and SCR), which are both co-adapted and have extraordinarily high polymorphism. Sequences of both genes will be obtained from a sample of plants collected from a natural population in Iceland, together with analysis of hand-pollination data to determine the plants' incompatibility groups. The results will test co-segregation of SRK sequences and incompatibility alleles, and will provide samples of alleles large enough for testing for linkage disequilibrium between the pistil and pollen incompatibility loci, and to test for recombination and balancing selection.

C.27 A large-scale comparative study of microsatellites in microbial genomes: the role of repetitive DNA in hypermutability, generation of phenotype, and adaptation

Dr D Field.

The wealth of genome sequences currently available provides us with an unprecedented opportunity to study the molecular basis of phenotype and the ecological significance of genetic variation. Despite their mutability, potential utility as molecular markers, and role in phenotypic evolution, microsatellites remain a largely unexplored feature of non-eukaryotic genomes. We propose to conduct a comparative genomic study of the genomes of bacteria, viruses, plasmids, and organelles aimed at identifying factors that control repeat evolution. We will also collect population-level data on repeats in H. influenzae that control the expression of genes known to mediate interactions with the environment. Our work should provide new perspectives on comparative genomic architecture, stability, and the molecular mechanisms underlying phenotypic evolution.

3.1.2 Identifying the mechanisms of sexual disruption in fish exposed to endocrine disrupting chemicals and their mixtures using transcriptomics.

It is now well established that chemicals in effluents from sewage treatment works are oestrogenic to fish and cause disruptions in sexual development (intersex; the simultaneous presence of both male and female sex cells in a single gonad) in wild, indigenous fish populations [see Case Studies; Tyler *et al.,* 1998; Purdom *et al.,* 1994; Jobling *et al.,* 1998; Jobling *et al*., 2001). New genomic approaches offer model-independent means of defining the pathways of EDC action and by identifying the affected genes and proteins can direct the development of systematic screens for EDCs. A project funded in the NERC Environmental Genomics Thematic to the Universities of Exeter and Liverpool is applying genome-level (transcriptomic) analyses, together with assessment of whole organism responses to identify genes whose regulation during normal gonadal differentiation is altered as a consequence of exposure to individual EDCs and their mixtures during early life in the roach (Rutilus rutilus). Roach are the principal study species in this project because intersex is widespread in roach populations living in effluent contaminated rivers in the UK and there is a detailed histological understanding of its gonad development. There are, however, no genomic resources for the roach thus the analysis of roach is being complimented with studies of the zebrafish (ZF). The zebrafish is one of the most widely used fish species in developmental biology and ecotoxicology (including for studies on EDCs) and is one of the key species used by international regulatory authorities (e.g. OECD) in standardised tests (Huet 2000). In addition to its well-appreciated experimental conveniences, ZF already has large-scale expression sequence tag (EST) clone collections, and in 2003 it's complete genome sequence will be available. The zebrafish is a member of the Cyprinidae (the largest family of freshwater fishes globally and in the UK) and the hormonal processes and mechanisms controlling reproductive development established so far are very similar to that of our native fish species. This work will assess whether the cascades governing gonad sex cell and duct disruptions in cyprinid fish are generic. In turn, the project will assess the applicability of the zebrafish for informing potential impacts of EDCs on wild, indigenous, cyprinid fish.

In the first phase of the project gonadal cDNA clone sets and cDNA microarrays are being established to assess gonad sex differentiation in roach and zebrafish. Given that there are currently no genomic resources for roach, an important outcome of this project is to create full-length, directionally cloned gonad cDNA libraries. For zebrafish, ~10,000 clones will be selected containing relevant candidate genes (StAR, vasa, nanos-1, DMRT, DMO, aromatases and various other steroidogenic enzymes, oestrogen receptors α, β & γ, Sox genes etc.). The project will then identify gene pathways of disruption of gonadal differentiation in roach and zebrafish exposed to individual EDCs. In this phase embryos/early life stages will be exposed to specific EDCs, namely, ethinyloestradiol (EE2) and 4 nonylphenol (4 NP), chemicals that are present in effluent discharges from STWs that have feminising effects in fish (in both roach and ZF), inducing vitellogenin synthesis and duct disruption at environmentally relevant concentrations (Toft *et al.,* 2001; Jobling *et* al., 1996; Christiansen *et* al., 1998; Thorpe *et* al., 2000; van Aerle *et al.,* 2002), and germ cell disruption (oocytes in the testis of 'males') at pharmacological doses (Piferrer and Donaldson 1992). The exposures to these EDCs will run throughout the period of gonad sex differentiation, and the fish will be sampled for gonad transcriptome and cellular and tissue level analyses at key stages in the development process for germ cell and gonad duct formation Rodgers-Gray *et al.,* 2001; Maack *et al*., 2002).

Having defined specific gene responses to known EDCs, a minimal set of genes that can distinguish gender-specific EDC-type effects will be defined. The performance of this set will be validates against STWs effluents. The STW effluents chosen for study are known to induce feminising effects in wild fish, including roach. In this phase of the project early life stage roach (0-100dph) will be exposed to a series of doses of effluent and sampled at key stages and gonad transcriptome and cellular and tissue level analyses undertaken.

In this project well-established biochemical and histological techniques will be employed to measure cellular and tissue level endocrine and gonad responses to the specific EDCs and their mixtures. Molecular analyses will involve the application of cutting edge technologies both to obtain progenitor gonad cells and gonadal duct cells from the early life stage and juvenile fish for RNA extraction (using laser capture microdissection of freshly dehydrated transverse histological sections) and to conduct the subsequent transcriptome analyses. Gonadal tissue collected for RNA extraction will include both germ cells and somatic cells as the transcriptomic responses of the whole gonad in effecting gonadal sexual differentiation will be assessed.

APPENDIX D BBSRC FUNDED POST-GENOMICS PROGRAMMES

Information on all funded programmes is available via the BBSRC Oasis Database.

http://dataserv.bbsrc.ac.uk/Welcome.html

Use the keyword search facility and type in the keywords ‘genomics’ or ‘functional genomics’.