Genes and the environment

Environmental Genomics programme 2000-2007
Environmental genomics

The specific field of environmental genomics answers questions such as:

- which genes are important for survival and reproduction?
- how do plants, animals and microbes respond to abrupt or temporary changes, for example, seasonal temperatures, or changing salt concentrations in tidal rivers or estuaries?
- why do some pollutants interfere with sexual development in fish populations?
- how did the genome itself evolve?

DNA

DNA – deoxyribonucleic acid – is the molecule that carries the genetic information needed to produce an organism and make it work. The information is carried in the precise sequence of four bases (Adenine, Thymine, Guanine and Cytosine) of which it is comprised.

Genes

Genes are units of genetic information and are made of DNA. Some contain the instructions for producing proteins, whilst others control and regulate the activities of other genes. Genes determine the inherited characteristics that distinguish one individual from another. Each human has an estimated 30-45,000 genes.

Genome

The complete genetic material of an organism including all the genes as well as additional non-coding DNA.

Genomics

Researchers have identified the entire DNA sequence of a growing number of species, and we are now developing ever greater insight into how the genome functions to give rise to a complete organism. This new capacity has created the field of genomics and has revolutionised biological investigation. Genomics offers new opportunities for investigation into how the genome interacts with the environment, for example:

- researchers can identify how tens of thousands of genes respond, at any one time, to environmental change. Until recently, they could only study one gene at a time.
- scientists can say how one species is related to another on a genetic level, markedly increasing our knowledge of evolution.

Environmental genomics

The specific field of environmental genomics answers questions such as:

- which genes are important for survival and reproduction?
- how do plants, animals and microbes respond to abrupt or temporary changes, for example, seasonal temperatures, or changing salt concentrations in tidal rivers or estuaries?
- why do some pollutants interfere with sexual development in fish populations?
- how did the genome itself evolve?
A genome-level approach to biological investigation has resulted in what many are now referring to as a ‘post-genomic era’ in biology, with high expectations for significant applications in biomedical and agricultural areas.

Indeed, scientists have produced complete DNA sequences for disease organisms. This allows researchers to identify specific mechanisms of virulence, leading to better prevention and treatment methods. Knowledge of how genes respond to the environment has enabled researchers to improve crops, for example, by creating drought resistant strains.

Similarly, the emergence of genomics tools for scientific investigation has presented new opportunities for environmental research. In the recent past, genetic studies were dominated by the notion that there was a one-way flow of information from genome to phenotype (the observable characteristics of an organism) to interaction with the environment. New methods and insights, many from this programme, have shown a greater interaction between genome-level processes (such as genome organisation and patterns of gene expression) and the environment than previously suspected.

Why do we need environmental genomics?
The principle aim of the NERC Environmental Genomics programme has been to use genomic tools to increase knowledge of the interaction between genome-level processes and the environment.

Prior to this programme, investigations of interactions between an organism’s genome and its environment involved detailed studies of the fitness consequences of one or a few specific genes at a time, or statistical analyses of quantitative variation that made assumptions about underlying genetic effects but made no attempts to measure genetic effects directly. One of the first applications of genomic methods was to develop detailed genetic maps of organisms in an attempt to close the gap between single gene studies and quantitative studies. This involves mapping the regions of the genome that influence quantitative variation, for example, height, life-span, age of sexual maturation or flower dimensions.
How does environmental genomics enhance environmental research?

Advances in understanding of genome structure have challenged the notion that an organism’s adaptation strategies, even for very specific environmental effects, can be traced to single genes. Instead, it has become clear that adaptation is due to integration across a range of genomic effects, including multiple genes and variation in gene expression patterns. Thus, there has been a growing mandate for a comprehensive genomic approach to understanding how organisms respond to their environment.

The Environmental Genomics programme focused on genomic methods that allow such a comprehensive approach. Throughout the programme, researchers applied such genomic methods to a wide range of environmental issues. These have included mechanisms that contribute to the origins of biodiversity, evolutionary adaptations that contribute to fitness, and organismic response to specific environmental challenges or changing environmental conditions. These methods have also provided new insights into the response to very specific biological and environmental challenges, such as disease or pollution.

Building solid foundations

As a result of this programme, we now have a solid foundation for future application of genomic technologies to increase knowledge of how organisms interact with their environment. The various projects funded by this initiative have demonstrated the feasibility and usefulness of genomic approaches to environmental sciences. The postgraduate students and post-doctorates who have been trained in these new technologies, have set the stage for future development of environmental genomics as a long-term component of the NERC research portfolio.
Key findings

- **Origins of biodiversity – molecular basis of adaptation.** Adaptations to specific environmental conditions often involve complex genetic mechanisms. Results from this programme have advanced our understanding of such complex adaptations as reproductive timing in Soay sheep, flowering time in relation to climatic variation in thale cress, and timing and duration of dormancy in nematodes. This information will inform conservation efforts in relation to such environmental challenges as habitat fragmentation and climate change.

- **Origins of biodiversity – genomic basis of adaptation.** A revelation of the genomics era is the role of genome organisation in adaptation and the potential for a direct genome-level role in adaptation. Findings from this programme have shown that: the rate at which different combinations of genes in the genomes can be generated through recombination varies in a manner that can regulate rates of adaptation; repetitive DNA fragments that are capable of ‘jumping’ around in the genome play a role in adaptation; and that the structure of the genome plays a significant role in establishing new species.

- **Environmental health – pollution.** Current approaches to addressing the consequences of pollution include determining the consequences of specific toxic chemicals on ecosystem function. The Environmental Genomics programme funded projects to assess the impact of particular chemical pollutants. One study addressed the effects of oestrogenic compounds in aquatic pollution that modify the sex of fish, which in turn has broad implications for conservation of wild species as well as human health. Another study examined the biological impacts of terrestrial pollutants, such as agrochemicals, cadmium and aromatic hydrocarbons. Using earthworms as a ‘sentinal’ species, scientists determined the biological significance, and specific impacts, of varying levels of pollution.

---

NERC’s Environmental Genomics programme continues to have an enormous influence not only in the field of ecotoxicology but also in the basic understanding of receptors, modes of action of chemicals, and molecular background of toxicology. The studies on flounder and roach using microarray and ecotoxicology are well-known in the world.

Taisen Iguchi, National Institutes of Natural Sciences, Japan
The Environment Agency has benefited from the new community of researchers, established by NERC, and the data generated by this programme. The programme has provided high quality training and developed national expertise in a new field. Researchers from the programme developed beneficial dialogue between the research community and end users. The latter including both policy and industry regulated groups that need a sound understanding of genomic technologies and the implication for regulatory activities.

Dr Jim Wharfe, Head of Science Programmes, UK Environment Agency

**Environmental health – disease.**

Outbreaks of disease and the emergence of new diseases represent a major threat to natural ecosystems as well as to human health. The aim of environmental genomics approaches to disease include investigating how environmental conditions affect disease susceptibility in host organisms and how disease organisms respond at a genomic level to environmental change. Environmental genomics researchers found that different suites of genes are expressed in organisms subject to environmental extremes of temperature or parasitoid load in fruit flies, a widely used model organism. Environmental Genomics researchers also investigated the genetics of the tick-borne Great Island virus that infects seabirds. Using 399 isolates of this virus from the UK, Norway and North America, they found that strains sampled from seabird colonies around the UK exhibit little genetic difference but, like the influenza virus, different strains regularly swap whole genes.

**Genes and Ecosystem Services.**

The natural environment provides a range of ecosystem services, including food, medicine, fuel, clothes, timber, climate regulation, water purification, soil regeneration, nutrient cycling, waste recirculation and crop pollination. Many ecosystem services are provided by bacteria in the environment, ranging from decomposition, breakdown of toxic pollutants, and nutrient cycling. Using genomic tools, environmental genomics researchers have identified specific groups of genes involved in such services, making it possible to better manage microbial impacts on ecosystem services.
When faced with an environmental challenge, for example changes in salinity or temperature, organisms can respond in several different ways. They can evolve through the spread of new variants of genes, so that they are better able to cope with the challenge. This process is called adaptation through natural selection. Alternatively, some organisms can change their physiology, behaviour or life cycle to meet the challenge, but without changing their genes. This is called acclimation or plasticity. If an organism is unable to do either, and the challenge is severe enough, the organism may become locally extinct.

As the world is faced with increasing environmental change, such as global warming, pollution and habitat destruction, we need to know more about the molecular basis of adaptation, and the relationship between adaptation and acclimation.

**Genetic solutions to environmental change**

Researchers have made great progress recently in identifying the genes that control continuously varying features, such as disease susceptibility or the timing of life history stages such as flowering or reproductive age. These genes, or regions of the genome, are known as Quantitative Trait Loci (QTL) (see box opposite).

Scientists have mostly studied these type of genes in the laboratory with organisms where it is possible to make known crosses in controlled environments. The Environmental Genomics programme showed that it is possible to study these genes in the wild. The team used pedigree information in a well-studied population of wild Soay sheep to identify QTL for body size, reproductive timing and parasite resistance. This is useful because it shows that we can track the genetic changes that take place during the adaptation of organisms to their changing environment.

**Genetic response to climate change**

Environmental responses at the limits of species distribution provide insight into how species might respond in the future to climate change. Plants in the UK now flower on average 4.5 days earlier than they did four decades ago and this change has been attributed to global warming.

Scientists studied the impacts of temperature on flowering time in thale cress (*Arabidopsis thaliana*), a close relative of many crop species, from the Arctic to temperate climates. Thale cress is found throughout Eurasia, and getting flowering time right is critical to the plant’s survival. Too early in the Arctic, and the plant will freeze; too late in the Mediterranean and the plant will wither before flowering is complete. The length of cold hardening has a marked effect on the flowering time.

Genomic tools developed for this species made it possible for scientists to locate the specific gene that regulates flowering time and how this gene works in combination with other genes that mediate environmental response. Temperature response is a critical feature of adaptation and will contribute to conservation management of other species as they respond to climate change. The scientists found that most variation was associated with a single gene. Now that this gene and its effect have been discovered in thale cress, genomic tools can be used to investigate its impacts on flowering time in other species.
To be or not to be dormant
For the free-living nematode worm (Caenorhabditis elegans) juvenile stages sense their environment and, if conditions are good, that is sufficient food is available, then they develop into reproductive adults. However, if conditions are not good, that is there is insufficient food available, then they become dormant and do not develop into reproductive adults. The larvae only resume development to the adult stage when conditions once again become good.

So, depending on the quality of the environment, the nematode may or may not become dormant. This is a classic example of an acclimation response to environmental change. The genetic mechanisms of how this occurs are well known, but Environmental Genomics scientists studied different strains of the roundworms that show different abilities to form this dormant stage in response to environmental change. They found that different strains show differences in fitness, and that the differences are associated with two new QTLs, which control whether the genetic system that produces dormancy is expressed or not. This shows that the genes for adaptation are not necessarily the same as the genes responsible for acclimation.

Fur and feathers: same genes
As in the example of the roundworm, classical and molecular genetics have identified many genes that are involved in particular traits. It is important to know whether any or all of these genes, ‘candidate genes’, are actually involved in the genetic changes that lead to adaptation. One study in the Environmental Genomics programme investigated variation in the plumage colour of birds and found that some of the genes giving different feather colours in chickens and other birds are the same as genes giving variation in coat colour in mammals.

Genes that allow metals to accumulate in plants
Another study investigated a phenomenon in plants, hyperaccumulation, where plants can accumulate hundreds of times more metal in their leaves than is normal. It is not known exactly why plants accumulate metals, but one possibility is that it protects plants from pests and diseases, while another is that the plants accumulate the metal while taking up another nutrient. Physiological and genomic studies identified a number of candidate genes, but analysis showed that most of these are not responsible for the adaptation since they are not QTLs.

Ancient origins of internal clocks
Environmental Genomics scientists studying how ragworms (Nereis virens) manage their internal clocks have found that they possess many genes that are the same as clock genes in other organisms as diverse as fungi, insects and humans. Terrestrial organisms have clocks mostly based on a 24-hour cycle (circadian clocks), but the tidal ragworm has to have the ability to cycle in tandem with the tides as well. They have demonstrated that the ragworm shows both tidal and circadian clocks, and they suggest that they may both be variants on a common ancestral clock that evolved over 1.5 billion years ago. At this time the world rotated faster than now, so a day was only about 12 hours long – around the length of the current tidal cycle.
New species arise from variation in natural populations. In order to understand the origin of biodiversity, we need to understand how variation is generated in natural populations. A long-standing premise of evolutionary biology is that such variation is due to variation within a set framework of genes. However, one of the revelations of the genomic era is that genome structure and organisation, the distribution of genes along the genetic map, plays an important role in the generation of variation. Application of new genomic technologies makes the whole of genome structure a potential object of

Genetic solutions to encourage genetic diversity

During sexual reproduction, genetic diversity arises from recombination of genes along chromosomes as well as combining genes from different parents. In some plant species, genetic incompatibility mechanisms prevent individuals from mating with themselves because this causes loss of genetic diversity. Such species provide a useful basis for investigating both population level and genomic level processes that maintain and generate genetic diversity.

Self-incompatibility in the rockcress (*Arabidopsis lyrata*) involves two complementary genes close together on a chromosome that determine pollen and stigma compatibility. Reduced recombination between these genes is adaptive to ensure the appropriate gene combinations to prevent ‘selfing’.

Researchers investigated the amount of recombination in the vicinity of incompatibility genes in rockcress in order to determine how strongly this gene combination is maintained. Using genomic tools derived from thale cress, scientists identified DNA sequences found at varying distances along the chromosome away from this pair of genes. They found that recombination between DNA sequences in the vicinity of the incompatibility genes was lower than recombination between sequences located further away on the chromosome. This finding supports the hypothesis that natural selection has reduced the amount of recombination in this part of the genome, and helps us to understand why different genome regions evolve at different rates.

Genes that jump around the genome

New gene combinations arising from recombination is just one means by which genomes change over time. Genomes have proven to be remarkably
fluid in other ways. Transposable elements, pieces of DNA that move around within a genome, were discovered in the mid-20th century by Barbara McClintock, who received a Nobel Prize for her efforts. Since that time, it has been found that these ‘jumping genes’, or transposons, comprise a large percentage of the genome of most animals and plants; but the role of these ‘jumping genes’ in generating biodiversity and enabling evolutionary adaptation has remained a mystery.

Genomic tools provide the key to working out their role in adaptive evolution. Drawing on DNA sequence information contained in genomics databases, such as Genbank, and using the recently introduced method of quantitative PCR (see page 20), researchers measured variation in the number of copies of specific transposons in plant genomes to elucidate variation among individual plants of white campion (*Silene latifolia*) and related species. They found variation in transposable elements within a species to be related to ecologically relevant plant traits, for example, flower size.

**Genomics to conserve biodiversity**

Differentiation between species is key to the origin of biodiversity. In some plant groups, such as the many species of ragworts (*Senecio*), species retain the ability to hybridise, and hybrids themselves are potentially a source of new species. Indeed, hybridisation is one of the most important mechanisms of abrupt speciation in wild plants so has strategic importance to NERC in terms of population dynamics, establishing species in new habitats and conservation of biodiversity.

Even though some species are capable of hybridisation, the genomes of the parent species will have gone through sufficient independent evolution to undergo some degree of differentiation. Bringing two similar but divergent genomes together in a hybrid plant triggers genomic responses that might not be shown in either parent species. Using microarray technology, scientists investigated gene expression in the non-native diploid hybrid Oxford ragwort (*S. squalidus*), its parental species from Sicily (*S. aethnensis* and *S. chrysanthemifolius*) and the hybrid Welsh ragwort (*S. cambrensis*), a recent new species formed in the UK by hybridisation between Oxford ragwort and groundsel (*S. vulgaris*). These analyses revealed that large-scale genome-wide changes in patterns of gene expression accompanied both types of hybrid species formation and that such patterns were consistent between natural and resynthesised hybrids.

Population genetic models that have guided our view of evolutionary change over the past half-century remain valid, but they are only part of the story. Genomic investigations are providing additional insights into evolutionary processes that generate and renew biodiversity. Such insights will play a critical role in future efforts to conserve and protect biodiversity, that itself is a critical component of our sustainable future.
Waste is an inevitable product of society, pollution should not be. NERC’s strategy, Science for a Sustainable Future 2002-2007, set a challenge to develop tools and technologies to address more holistic methods of environmental risk assessment including identifying new and improved indicators of ecosystem health to monitor progress towards sustainability.

The scientific and regulatory efforts in the latter part of the 20th century sought to identify which chemicals cause damage to ecosystems so that their use could be restricted or prevented. In the early part of the 21st century environmental genomics is unravelling the ecologically important mechanisms of toxicity and determining why some pollutants are more toxic than others. Environmental genomics provides tools to determine the health of wildlife populations and identify susceptible individuals within natural populations to assist many aspects of conservation biology and environmental management. In the medium to longer-term researchers envisage these tools will reduce the numbers of animals, such as fish, used in regulatory assessments.

Genomic solutions to pollution
Current approaches to assess the potential impact that pollutants have on aquatic and terrestrial organisms rely on measuring whole-organism responses, for example, mortality, growth and reproduction. Whilst these approaches are useful for identifying chemical pollutants of potential concern, they provide little knowledge about the mechanisms of toxicity. Without the additional knowledge that can be gained through the integration of a genomic approach, it will be difficult to address some of the key scientific and regulatory challenges that currently face aquatic ecotoxicology. Such challenges include:

- predicting the influence of single and multiple exposures on the growth and sustainable development of wildlife populations.
- determining the impact of low dose exposures on the health of wildlife populations, including chemical cocktails or mixtures such as sewage effluents.
- predicting toxicant responses across a very broad diversity of wildlife populations and estimating how changes within one trophic level will affect ecosystem structure, for example predicting population-level effects.

In order to meet these challenges the Environmental Genomics programme funded projects to assess the impact of chemical pollutants on aquatic and terrestrial organisms that engaged relevant industrial and regulatory stakeholders. The projects also evaluated the applicability of model organisms, such as zebrafish, where we already have a wealth of genomic data, against indigenous UK species for which only a limited genomic resource existed.
Chemical pollution and the sex of fish

Chemicals contained within effluents from sewage treatment works discharging into UK rivers are known to be oestrogenic and cause feminising effects in wild male fish. Similar effects in other wildlife species throughout the world, and links with altered reproductive health in humans, have caused international concern and led to comprehensive screening and testing programmes for these endocrine disrupting chemicals (EDCs).

Scientists in the Environmental Genomics programme have developed extensive genetic and genomic resources, including gene arrays to study the expression of thousands of genes simultaneously. This work has focused on two species of fish in particular, namely, the roach (Rutilus rutilus, found to be sexually disrupted in UK rivers) and zebrafish (Danio rerio, a species used extensively for the regulatory assessment of chemicals).

These tools have identified the genes disrupted as a consequence of exposure to individual EDCs (nonyl phenol and the human contraceptive ethinlyoestradiol) and mixtures (a treated sewage effluent). Long-term exposures (greater than two years) to ethinlyoestradiol and treated sewage effluent induced all of the feminised responses seen in wild fish. The team identified a suite of genes key to the process of sexual differentiation whose normal pattern of expression was altered by ethinlyoestradiol.

These are amongst the longest and most comprehensive studies conducted to date for both the zebrafish and roach. This is the first study to demonstrate:

- chemical pollutants in domestic sewage effluents induce true intersex in exposed fish.
- exposure to levels of ethinlyoestradiol already found in the environment induces complete gender re-assignment in roach.
- that exposure to ethinlyoestradiol during early life induces intersex in fish that appears to pre-sensitise females to oestrogen in later life.
- a suite of potential biomarker procedures that researchers could use to assess whether chemicals can adversely affect the sexual development of fish, for example, 42 potential biomarkers in roach.

‘Big worm, little worm’ – genes to populations

Increasingly researchers are harnessing the discriminatory power of genetic techniques to produce individual clinical diagnosis, for example, personalised medicine. Scientists are transferring these same principles to the field of environmental diagnostics, where site-specific damage caused by chemical pollutants is a major consideration in terms of ‘fit-for-use’ decision-making, agricultural sustainability, wealth creation, and human as well as environmental health.

The concept is appealing but is constrained by our limited knowledge of the fundamental biology and genetics of most soil dwelling animals. To overcome the difficulty scientists in the Environmental Genomics programme focused on a widely distributed earthworm species capable of living in pristine soils, as well as in polluted brownfield and industrial soils. These ‘ecological engineers’, praised so highly by Charles Darwin, are potentially faithful reporters of the quality of the soils in which they live.

The team examined the earthworm genome when exposed to an agrochemical (atrazine), a heavy metal (cadmium) and a polycyclic aromatic hydrocarbon (PAH; fluoranthene). They compared the earthworm’s genetic-fingerprints with the molecular genetic responses to the same toxic
Environmental genomics research has sequenced and annotated around 17,000 genes for the earthworm. Environmental genomics research has sequenced and annotated around 17,000 genes for the earthworm and generated data on how the earthworm responds on a genetic level to specific pollutants, together with all the genes for the model roundworm (*C. elegans*). The team integrated the response patterns for both species in order to understand the activities of different gene pathways involved in development, seasonal fluctuations, and toxicant exposure. The researchers complemented this genetic data with studies on the protein composition of both organisms to provide an insight into the mechanism of response that these representative terrestrial organisms have to chemical exposure. The programme team have assembled all this information, together with a burgeoning wealth of genetic data on other key terrestrial and marine annelid worms, in a public-access database called LumbriBASE see: www.earthworms.org.

They also used genetic manipulation to produce a strain of (*C. elegans*) that bio-fluoresce (they light up) when exposed to pollutants, specifically heavy metals, such as cadmium. These fluorescent worms can be used to assess the extent to which soils contain metal pollutants.

**National and international leadership**

NERC research into environmental toxicology is truly world-leading and has major stakeholder interest from the chemical and water industries and environment protection groups. The programme funded several knowledge transfer workshops and networks to examine the role that genomics can play in environmental toxicology. These workshops involved close links with with the Environment Agency, the Society of Environmental Toxicology and Chemistry (SETAC) and other UK stakeholders as well as government bodies from the United States, Canada and Japan.

The relationship between the genome and the environment, as applied to environmental risks and hazards of pollution and wastes, is a priority area of science within the NERC strategic plan and genomic approaches will continue to help us to improve the scientific basis for identifying and understanding the risks that chemicals pose to wildlife.
Outbreaks of disease and the emergence of new diseases represent a major threat to natural ecosystems as well as to human health. For example, the spread of plant diseases such as chestnut blight and Dutch elm disease have caused landscape-level changes in the composition of tree species, and newly introduced diseases in the UK, such as sudden oak death, have the potential to effect changes on a similar scale.

Emerging new diseases such as HIV, SARS, and the current concern over the possibility of an influenza pandemic are an ongoing threat to human health. Major drivers of disease spread or emergence include environmental conditions, in particular climatic fluctuations or human-caused disturbance, and genetic changes in the disease organisms themselves.

The aim of environmental genomics approaches to disease include investigating how environmental conditions affect disease susceptibility in host organisms and how disease organisms respond at a genomic level to environmental change.

Changing environments present serious challenges to all organisms in terms of stress and its impact on disease resistance. Genomic tools provide a means of understanding the genetic basis for stress response. Working with the geneticist’s fruit fly (Drosophila), NERC-funded scientists used microarray technology to survey which genes are activated during stress response and to compare the expression profiles of strains that differed genetically in their ability to deal with environmental challenges, including temperature and attack by parasitoids – insects whose larvae develop by feeding on the bodies of other insects.

Temperature is a major determinant of where an organism can live, and organisms that do not have physiological temperature regulation are particularly sensitive to temperature extremes and also to long-term exposure to mild thermal stress. Parasitoids also represent a serious environmental challenge to their hosts.

Genes for environmental extremes
Using microarray technology, researchers found that different suites of genes are expressed in organisms subject to environmental extremes of temperature or parasitoid load. Moreover, we found that both fruit flies subjected to laboratory-based selection for resistance to such environmental extremes and resistant fruit fly strains express different suites of genes than non-resistant strains. Thus, environmental stress, including pathogens, induces a genome level response within the lifespan of a single organism as well as causing genomic changes over time through natural selection.
Evolutionary arms race
The evolution of resistance may also vary from one population to the next, especially in the case of diseases where the causative organism shows genetic variation. Such a phenomenon can lead to an ‘arms race’ with the evolution of resistance in host organisms responding to ongoing changes in the virulence of disease organisms. Viruses, in particular, evolve very rapidly. NERC scientists investigated whether natural populations of wild relatives of cabbage (Brassica oleracea) respond differently to attack by local strains of turnip mosaic viruses than to viruses from a distant locality. If they do, what genes in the plants are responsible for such differences? Researchers obtained plants and local viruses from two well-separated coastal regions of Britain, North Wales and Dorset. The team found that local populations and their associated strains of virus have differentiated, so that introducing the disease from one population to another results in different levels of infection and host response.

Using microarray technology, they investigated patterns of gene expression in response to infection. They found differential gene expression between populations and species in response to the different viral types, and identified hundreds of genes that were responding to virus attack. They are currently looking for patterns that may enable them to identify the key genes involved and so stimulate future research.

Driving pathogen diversity
Many diseases have a complex life history involving multiple hosts, and are thus subject to a wide diversity of host organism environments. Tick-borne viruses that subsequently move on to other organisms through tick bites are an example. Genomic tools also have potential to provide a broader}

Environmental health

Disease

Evolutionary arms race
The evolution of resistance may also vary from one population to the next, especially in the case of diseases where the causative organism shows genetic variation. Such a phenomenon can lead to an ‘arms race’ with the evolution of resistance in host organisms responding to ongoing changes in the virulence of disease organisms. Viruses, in particular, evolve very rapidly. NERC scientists investigated whether natural populations of wild relatives of cabbage (Brassica oleracea) respond differently to attack by local strains of turnip mosaic viruses than to viruses from a distant locality. If they do, what genes in the plants are responsible for such differences? Researchers obtained plants and local viruses from two well-separated coastal regions of Britain, North Wales and Dorset. The team found that local populations and their associated strains of virus have differentiated, so that introducing the disease from one population to another results in different levels of infection and host response.

Using microarray technology, they investigated patterns of gene expression in response to infection. They found differential gene expression between populations and species in response to the different viral types, and identified hundreds of genes that were responding to virus attack. They are currently looking for patterns that may enable them to identify the key genes involved and so stimulate future research.

Driving pathogen diversity
Many diseases have a complex life history involving multiple hosts, and are thus subject to a wide diversity of host organism environments. Tick-borne viruses that subsequently move on to other organisms through tick bites are an example. Genomic tools also have potential to provide a broader
understanding of the fundamental processes that drive pathogen population diversity in the face of the sequence of environments they encounter over time.

NERC scientists addressed this process by investigating the genetic structure of the tick-borne Great Island virus that infects seabirds. They specifically designed the work to test whether antibody mediated immune selection of Great Island virus results from particular genes. Using 399 isolates of this virus from the UK, Norway and North America, they found that strains sampled from seabird colonies around the UK exhibit little genetic difference but, like the influenza virus, different strains regularly swap whole genes.

The two genes that form the outer surface of the virus particle had a high proportion of nucleotide mutations that cause changes to the protein sequence. Such changes to surface proteins of pathogens are usually caused by host antibodies and a field experiment showed that Great Island virus is indeed subject to intense antibody mediated immune selection. Despite this selection, and contrary to expectation, even the genes encoding the outer surface of the virus (and it's the outer surface that determines whether or not antibodies inactivate a virus) do not appear to form exclusive sets.

**Great Island virus is subject to intense antibody mediated immune selection.**
The natural environment provides us with food, medicine, fuel, fibres, timber, climate regulation, water purification, soil regeneration, nutrient cycling, waste recirculation and crop pollination. These are all ecosystem services.

Ecologists have long sought to understand the relationships between different ecosystem components; how factors such as biodiversity affect the maintenance of these services; and how they might be affected by environmental change. Understanding the latter has become urgent in view of escalating threats such as pollution, over-exploitation of fisheries, deforestation, erosion of soils, and increasing urban sprawl.

Genomics allows ecologists and evolutionary biologists to explore links between ecosystem structure and function, as well as generating practical tools, known as ‘genome-based sensors’, for directly monitoring changes in processes such as recycling waste materials. Activity is especially focused on microbes, and identifying the major genes and species determining such transitions and conversions.

**Recycling nutrients, nitrogen fixing and the Coca-Cola hypothesis**

Dead and decomposing organisms provide a crucial ecosystem service: recycling nutrients – including the cycling of nitrogen, carbon and phosphorus, the building blocks of many species. Bacteria have long been known to control such cycles, though before the advent of genomics it was difficult to investigate the genes controlling such activities because scientists could only rarely culture them in the laboratory.

Environmental Genomics scientists have now been able to examine millions of genes from uncultured, unknown soil bacteria. By sequencing these genes they found many new species of bacteria, and by comparing these unique genes with those already described, it will be possible to examine the contributions of particular species to the recycling process.

Intriguingly, it was not possible to identify the genes involved in the ‘fixing’ of biological nitrogen, that is, a process in which gaseous nitrogen is converted by bacteria into a form available for plant growth. It seems that the bacteria containing the genes for this process are much rarer in the soil than was previously thought – an idea summarised in what is called the ‘Coca-Cola hypothesis’. Only a tiny proportion of the working population is directly involved in manufacturing soft drinks, but those that do, can do so very efficiently. Therefore the beverage is present in cafes, bars and shops from Malta to Mongolia. Similarly, only very few cells, perhaps one in a thousand, in the overall bacterial population may fix nitrogen, but these specialists are also highly efficient at their jobs, therefore ensuring ecosystem and global distributions.

Although they did not find these specific genes, the hunt continues. The gene libraries are immortalised in freezers so that they can be used by scientists to look for other genes in the future.

**Enhancing natural variation: using microbes to breakdown pollutants**

One particularly exciting contribution of genomics is based on the ability of many microbial populations, such as bacteria, to remove man-made contaminants from the environment, and restore a healthy ecosystem. Researchers have studied the role of a specific protein that regulates a wide range of bacterial activities, including the breakdown of chemicals, allowing their absorption and use by plants. By adopting a technology that allowed the simultaneous comparison of genes controlling the protein, it was possible to
It was possible to examine how bacteria performed in hundreds of different environments at the same time.

examine how bacteria performed in hundreds of different environments at the same time, allowing scientists to make predictions about the best conditions for bacterial growth.

The rapid processing of samples and analysis of large datasets has enabled them to identify new functions and give them an improved understanding of how to remove pollutants quickly from the soil. They can use this information to enhance natural variation and select bacteria with improved performance for the breakdown of pollutants, thereby increasing the efficiency and cost-effectiveness of ecosystem recovery.

Protecting plants from disease and fungi

The team found that the same protein used to regulate chemical breakdown also controlled the incidence of diseases in plants – an unexpected finding revealed only by the ability of genomics to identify and focus on the activity of specific genes, often in bacteria that cannot be cultivated in the laboratory.

Additional work has helped to overcome the difficulties associated with working with bacterial species that may at the same time be both beneficial and disease causing. Using a system of DNA tagging to identify new bacterial varieties, it was possible to isolate only those forms with reduced harmful effects and improved beneficial consequences. Varieties that were good both at breaking down pollutants and also in protecting the roots of seedlings of crop plants from harmful fungi and worms could be propagated for addition to soils.

Identifying environmentally useful bacteria

In terms of global development, harnessing the beneficial properties of these bacteria has the potential to replace the use of chemical pesticides and fertilisers, and maintain a clean environment in the face of increasing pollution. An associated studentship focused on developing more efficient methods to identify environmentally friendly bacteria and their associated genes for surviving in contaminated environments. The research allowed a greater understanding of the diversity of bacterial types, as well as the diversity of genes they use to mediate ecologically beneficial processes in the environment.

Marine viruses and genes that slow down the aging process

The oceans have long been known to have a dramatic influence on climate. One way they do this is by absorbing large quantities of the greenhouse gas, carbon dioxide. This process is mainly carried out by small, floating marine algae – phytoplankton. Among the various factors that influence the abundance of these important ‘carbon sinks’ is disease, including the effects of marine viruses. Environmental Genomics scientists studied a giant virus that infects chalk-covered algae (Emiliania huxleyi). These algae form beautiful oceanic blooms and soak up billions of tonnes of carbon dioxide.

Many of the genes in this virus have never been seen before, and some may slow down the aging process of the infected cell by keeping it healthy for as long as possible. A compound called ceramide is produced which can control the timing of cell death and may have applications in anti-cancer therapies and anti-aging creams in the cosmetic industry.

By comparing the structure of the virus with genes of known function from other organisms, we can better understand how the virus controls the growth of these ecologically important and highly abundant marine algae.
Data management and capacity

NERC’s Environmental Bioinformatics Centre

The Environmental Genomics programme provided an enviable legacy through its investment in the NERC Environmental Bioinformatics Centre (NEBC). Although aimed primarily at researchers working on the programme, the tools and materials provided by NEBC are used by scientists nationally and internationally. This investment in science and scientists has increased the capacity of the UK in environmental genomics and has raised the profile of NERC in this area.

The NEBC brings together data, computing and bioinformatics experts with a remit to provide tools and training, supporting the effective use of the new data types generated. This includes developing and providing tools for handling and processing data as well as supplying wide-ranging training and advice.

Each funded lab was provided with a powerful bioinformatics computer running Bio-Linux, an operating system and suite of easy-to-use bespoke and third-party bioinformatics software assembled by NEBC. The software is freely available for installation and there are more than 150 Bio-Linux systems running world-wide. A Live DVD version of Bio-Linux has also proved popular. It has been downloaded by people in the UK as well as from countries such as Poland, Mexico, Japan, Singapore, India, Taiwan, the United Arab Emirates, Canada and Brazil. They also provided teams with access to the commercial microarray analysis software system GeneSpring and GeneSpring Workgroup.

Investment in knowledge infrastructure is investment in our science and scientists

NEBC provides training in the form of courses, on-site visits to research labs and free access to all training materials. During the lifetime of the programme, researchers have had 35 training opportunities directly funded through the data management plan. This is bolstered by on-going support and collaboration opportunities through an online and phone helpdesk, attendance of research group meetings, one-to-one meetings and extensive web-based documentation.

Data created today has potential for tomorrow

Data sets can be re-used if they are described well enough for others to understand, and are made accessible. These conditions require effective data management throughout the research process. The scope of NEBC activities ensures that researchers have the necessary tools, data standards, guidelines and support to produce datasets for future use without an onerous investment of researcher time.

Proactive and co-ordinated data management is cost-effective and liberates researchers to do what they do best – good science.

NEBC is a centralised facility, with two university, research-based development partners. By acting centrally, the Environmental Genomics programme has sourced hardware and commercial software for much better prices than would have been possible otherwise. In addition, short-term access to bioinformatics and computing expertise is possible without draining researcher time, for example, by removing the necessity to write grants and hire staff for short-term bioinformatics tasks.
The life of an Environmental Genomics data set

Unlike those working in medical fields, many Environmental Genomic projects involved systems with little or no molecular data available before the project began. The team needed to make a large effort to generate information and experimental tools required to address scientific questions. The table below follows a typical scenario, highlighting where tools provided by NEBC were able to help along the way:

<table>
<thead>
<tr>
<th>Action</th>
<th>Data management at the NERC Environmental Bioinformatics Centre (NEBC)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record the project information. These include</td>
<td>Create an entry describing the project and planned outputs in the NEBC EnvBase data catalogue.</td>
</tr>
<tr>
<td>physical samples, digital data and journal articles.</td>
<td></td>
</tr>
<tr>
<td>Collect samples</td>
<td>Use NEBC HandleBar to barcode samples and track what has been collected and where it is stored.</td>
</tr>
<tr>
<td>Carry out DNA sequencing and submit to public repository</td>
<td>Use software, developed by NEBC partners in Edinburgh, to process and format a particular sequence type for submission to a public repository.</td>
</tr>
<tr>
<td>Create microarray and use for experiments</td>
<td></td>
</tr>
<tr>
<td>Analyse microarray data</td>
<td>Use tools such as Agilent’s GeneSpring, GeneSpring Workgroup, maxdView from NEBC partners in Manchester, or R/BioConductor. All are available through Bio-Linux.</td>
</tr>
<tr>
<td>Annotate microarray data and submit to a public repository</td>
<td>Use maxdLoad developed by NEBC partners in Manchester. This facilitates annotation to the internationally accepted MIAME standard and its environmental extension, MIAME/Env. Data can be exported for submission to the ArrayExpress microarray repository.</td>
</tr>
<tr>
<td>Update the EnvBase entry</td>
<td>Each existing holding is marked as verified if it has been created and is publicly available. Links are created to digital data, making all resulting data associated with a project easily locatable and accessible.</td>
</tr>
<tr>
<td>Discover and retrieve data sets for use in later research</td>
<td>Search the EnvBase catalogue, collect data via links to public repositories, including non-standard data types held at NEBC. For samples, contact the researcher, who can make use of barcodes to locate them.</td>
</tr>
</tbody>
</table>
DNA sequencing. Once a highly specialised and complex process, is now done regularly in many laboratories using automated methods. This system allows researchers to evaluate variation in DNA sequences among genes from different individuals. It also allows scientists to isolate and characterise different genes easily.

PCR (polymerase chain reaction) is a basic tool in genomics. It is likely that every single project involved in this programme used this technique. PCR involves taking a sample of DNA, placing it in a reaction mixture with the enzymes and raw chemicals needed for DNA replication, and running the mixture through an iterative series of replication cycles (heating to melt the DNA to single stranded, lowering the temperature to a suitable temperature for DNA replication to occur, and then cooling it to condense into double-stranded form again, repeat these steps about 40 times) in vitro to generate multiple copies of a target.

qPCR (quantitative PCR) is a relatively recent advance that extends basic PCR by quantifying the amount of double-stranded DNA at each replication cycle, enabling quantification of DNA copy number in a sample or levels of gene expression. Several environmental genomics projects used this method, and it is a tool that is being taken up more widely in genomic studies.

Quantitative trait loci (QTL) analysis is an approach to genetic mapping that has been made possible by genomic tools such as PCR. It allows us to simultaneously examine large numbers of genetic markers. These can be used to rapidly generate a genetic map for almost any organism. By extending such mapping exercises across large numbers of organisms in experimental or natural populations, scientists can statistically determine regions on the genetic map that contain genetic effects influencing complex characters, such as body size and other quantitatively varying traits influenced by multiple genes. By extending such mapping exercises across large numbers of organisms in experimental or natural populations, one can statistically determine regions on the genetic map, that is chromosomal regions, that contain genetic effects influencing complex characters, such as body size and other quantitatively varying traits influenced by multiple genes.

Microarray analysis is a powerful new genomics tool that integrates across a range of genomics technologies to enable simultaneous evaluation of expressions of thousands of genes. This technique was introduced only shortly before the inception of the Environmental Genomics programme, and enthusiasm over the ability to examine impacts of varying environmental conditions on entire suites of genes was a major motivation for many of the projects funded under this initiative.

Bioinformatics is the term applied to the use of sophisticated computational tools to investigate large datasets. The suite of statistical, computational, and data handling methods encompassed by this area are a key component of environmental genomics research.
Acknowledgements