

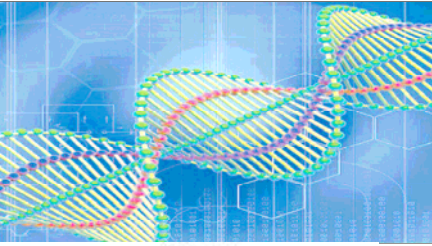
# Environmental Genomics: Setting the Scene

Dr Jason Snape  
Science Coordinator



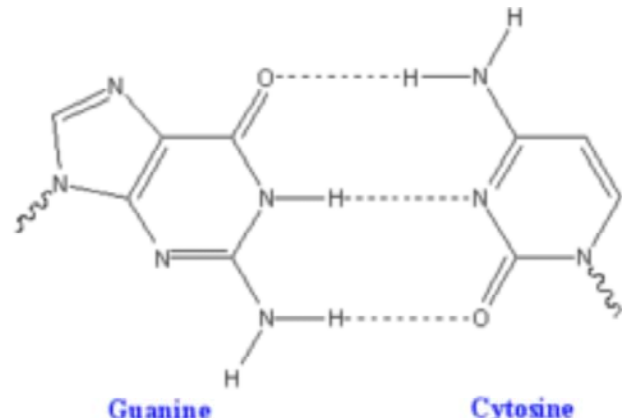
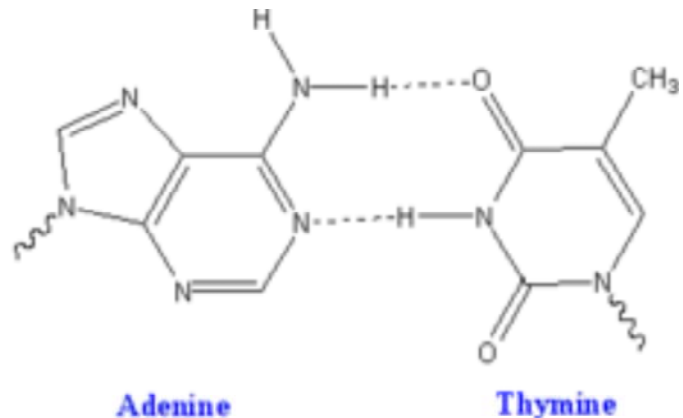
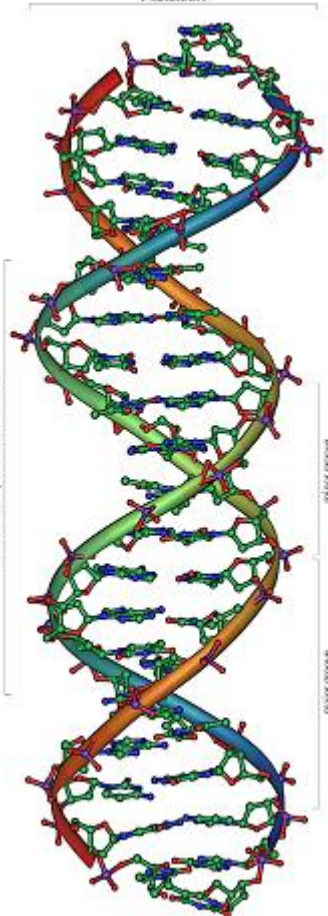
# Points to be Addressed

- **Genomics jargon buster**
  - terms
  - techniques/ technologies
- **Data management**
- **Science**
  - programme and scientific objectives
  - programme highlights



# Jargon Busting: Terms

**Deoxyribonucleic acid (DNA)** is a nucleic acid that contains the genetic information to produce and organism and make it work. This information is carried in the precise sequence of four bases of which it is comprised.





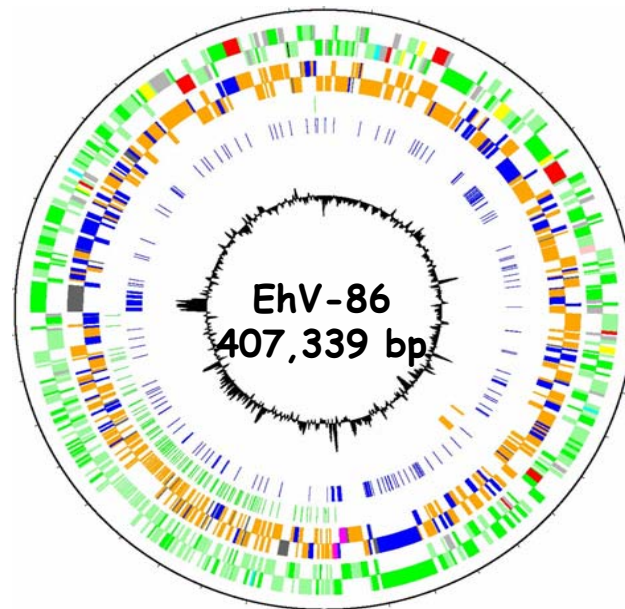
# Jargon Busting: Terms

A **Gene** is a segment of nucleic acid that contains the information necessary to produce a product, usually a protein. Genes correspond to units of inheritance. Some contain instructions for producing proteins, whilst others control and regulate the activities of other genes.



# Jargon Busting: Terms

The **genome** is often compared to a set of blueprints, since it contains the all of the genetic instructions to construct other components of the cell, such as proteins.



# Jargon Busting: Terms

## Gene content and genome size of various organisms:

### Species

### Size (Mb)

### No. of genes

*Mycoplasma genitalium*

0.58

500

*Streptococcus pneumoniae*

2.2

2300

*Escherichia coli*

4.6

4400

*Saccharomyces cerevisiae*

12

5800

*Arabidopsis thaliana*

125

25,500

*Caenorhabditis elegans*

97

19,000

*Drosophila melanogaster*

180

13,700

*Mus musculus*

2500

29,000

*Homo sapiens*

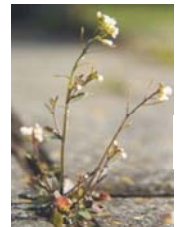
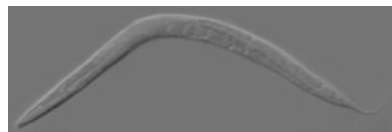
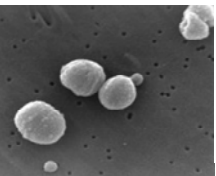
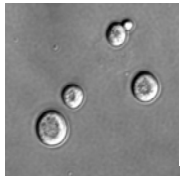
2900

27,000

*Oryza sativa*

466

45-55,000





# Jargon Busting: Terms

**Ribonucleic acid (RNA)** is another nucleic acid polymer consisting of nucleotide bases, that acts as a messenger between DNA and ribosomes that is responsible for making proteins out of amino acids.

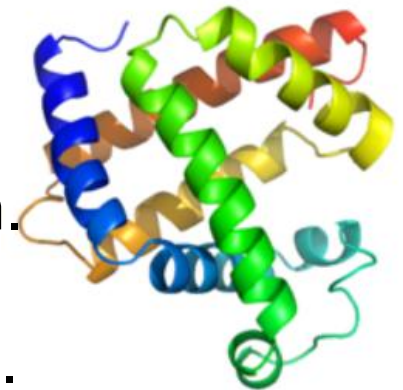
RNA serves as the template or transcript for the translation of genes into proteins.



# Jargon Busting: Terms

**Proteins** are essential parts of all living organisms and participate in every process within cells. They are made up of amino acids.

Many proteins are enzymes that catalyse biochemical reactions, and are vital to metabolism. Other proteins have structural or mechanical functions, such as the proteins in the cytoskeleton. Proteins are also important in cell signaling, immune responses, cell adhesion, and the cell cycle.

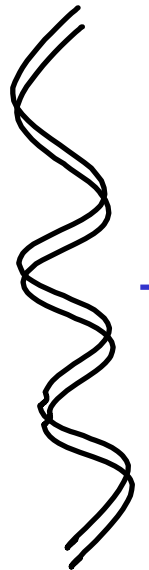


# Jargon Busting: Terms

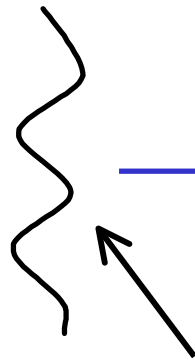
*Genomics*

*Proteomics*

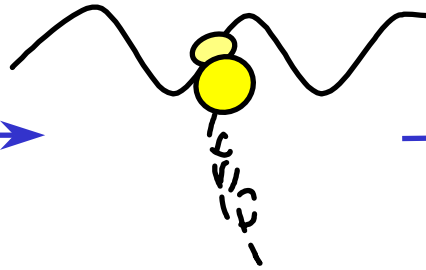
*Metabolomics*



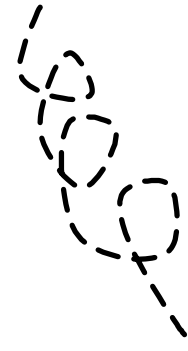
DNA



mRNA



Ribosome



Enzyme  $\geq 1$



Substrate

Product

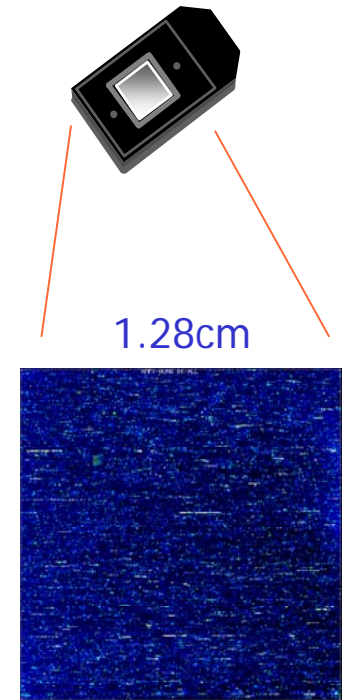
*Tools to measure at each level*



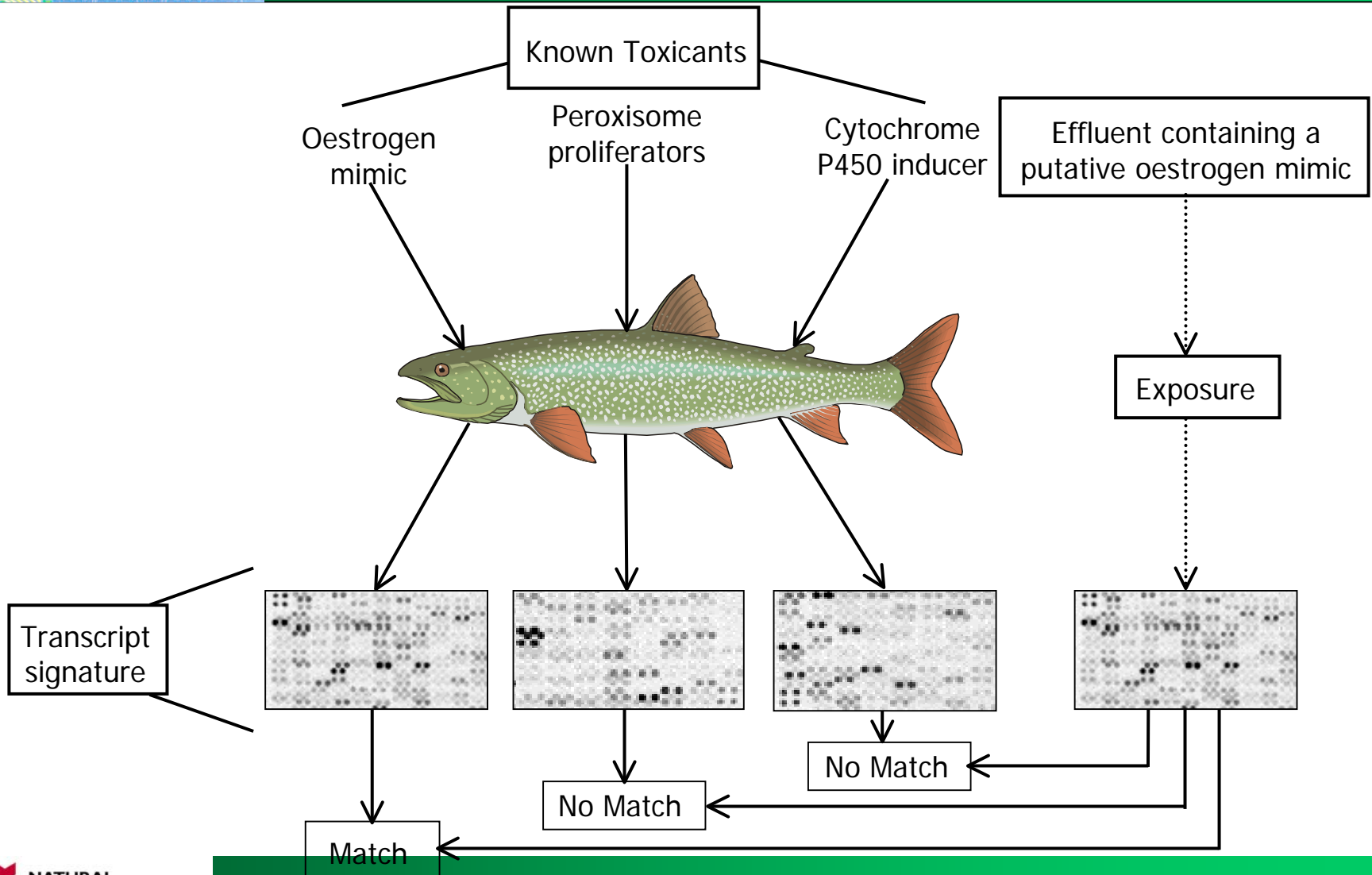
# Jargon Busting: Technologies

The **transcriptome** is the set of all messenger RNA (mRNA) molecules, or "transcripts", produced in one or a population of cells.

A **DNA microarray** (also commonly known as *gene or genome chip*, *DNA chip*, or *gene array*) is a collection of 1000's of microscopic DNA spots, commonly representing single genes, arrayed on a solid surface. The affixed DNA segments are known as *probes*



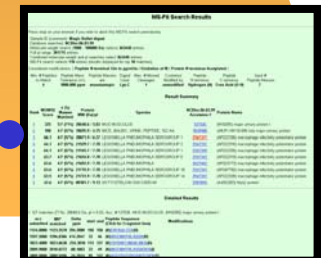
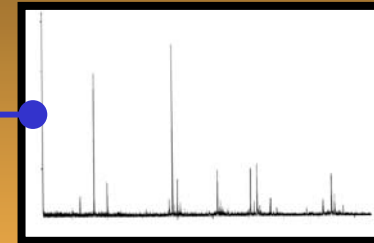
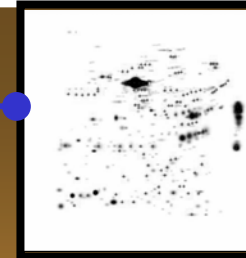
# Jargon Busting: Technologies



# Jargon Busting: Technologies

## Proteomics

- High resolution separation of proteins
  - 2-d electrophoresis
  - Spot excision
- Peptide fingerprinting
  - Proteolysis
  - Mass spectrometry
- Identification
  - Bioinformatics

A screenshot of a bioinformatics search results page, likely from a database like NCBI. It shows a table of search results with columns for accession numbers, protein names, and other identifiers. A blue line connects this image to the 'Identification' bullet point.

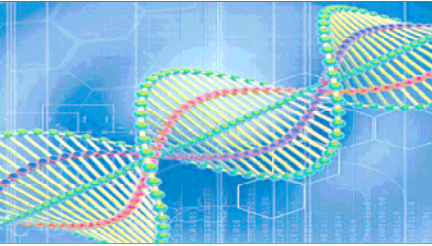
Accession	Protein Name	Species
U00096	lysS	Escherichia coli
U00096	lysR	Escherichia coli
U00096	lysX	Escherichia coli
U00096	lysZ	Escherichia coli
U00096	lysY	Escherichia coli
U00096	lysW	Escherichia coli
U00096	lysV	Escherichia coli
U00096	lysU	Escherichia coli
U00096	lysT	Escherichia coli
U00096	lysS	Escherichia coli



# Jargon Busting: Technologies

**Metabolomics** is the systematic study of the unique chemical fingerprints that specific cellular processes leave behind - specifically, the study of their small-molecule metabolite profiles. The metabolome represents the collection of all metabolites in a biological organism, which are the end products of its gene expression.

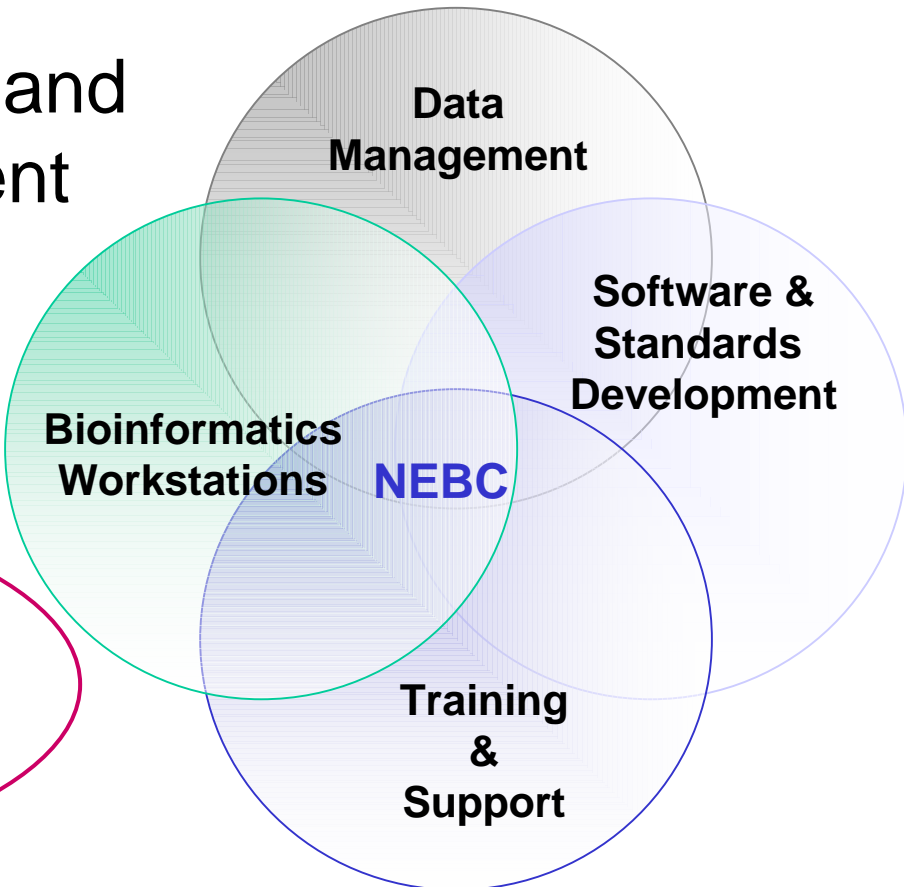
It is investigated using a combination of NMR and MS.

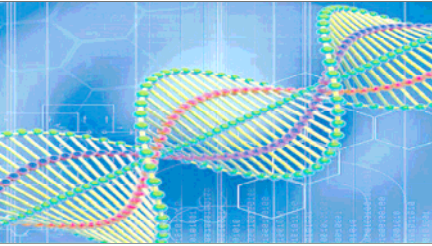


# Data Management

Environmental Genomics established NEBC to manage and coordinate its data management activities

- CEH Oxford, Manchester & Edinburgh





# Data Management



**Data management tools, skills, training & documentation have been delivered throughout the programme.**

**NEBC-developed tools that are available to the wider community. These include EnvBase, for cataloguing data and facilitating access, and HandleBar for barcoding, annotating and tracking physical samples.**

**Investment in data management tools and training means EG research outputs will be well annotated and accessible for public access and use.**

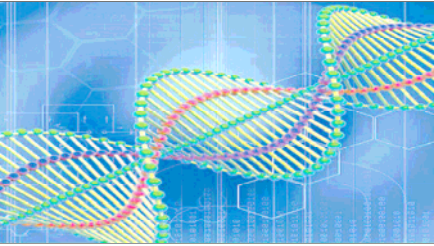




# Data Management

- Researcher time more effectively spent on research
  - automatic processing of sequences
  - custom scripts written to automate processing
  - analysis advice and collaborations provided
- Data quality improvements
  - Automated quality checking and cleaning of sequences
  - Microarray data quality evaluation provided
  - Tools and support encouraging higher quality annotation of data submitted to public resources (giving data a higher potential for re-use in the future.)

*Support continues until 2009*



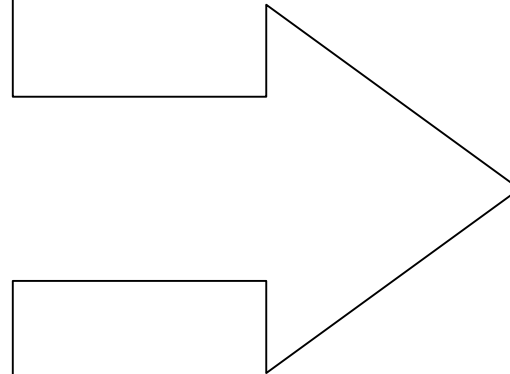
# Science – Top-down

## Environmental Genomics

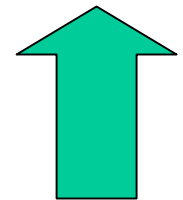
- SR2000 Allocation (c£17M)
- Capacity-building
- Stakeholder relevance

## Post-Genomics & Proteomics

- SR2002 Allocation (c11.6M)
- Targeted



CCGCC



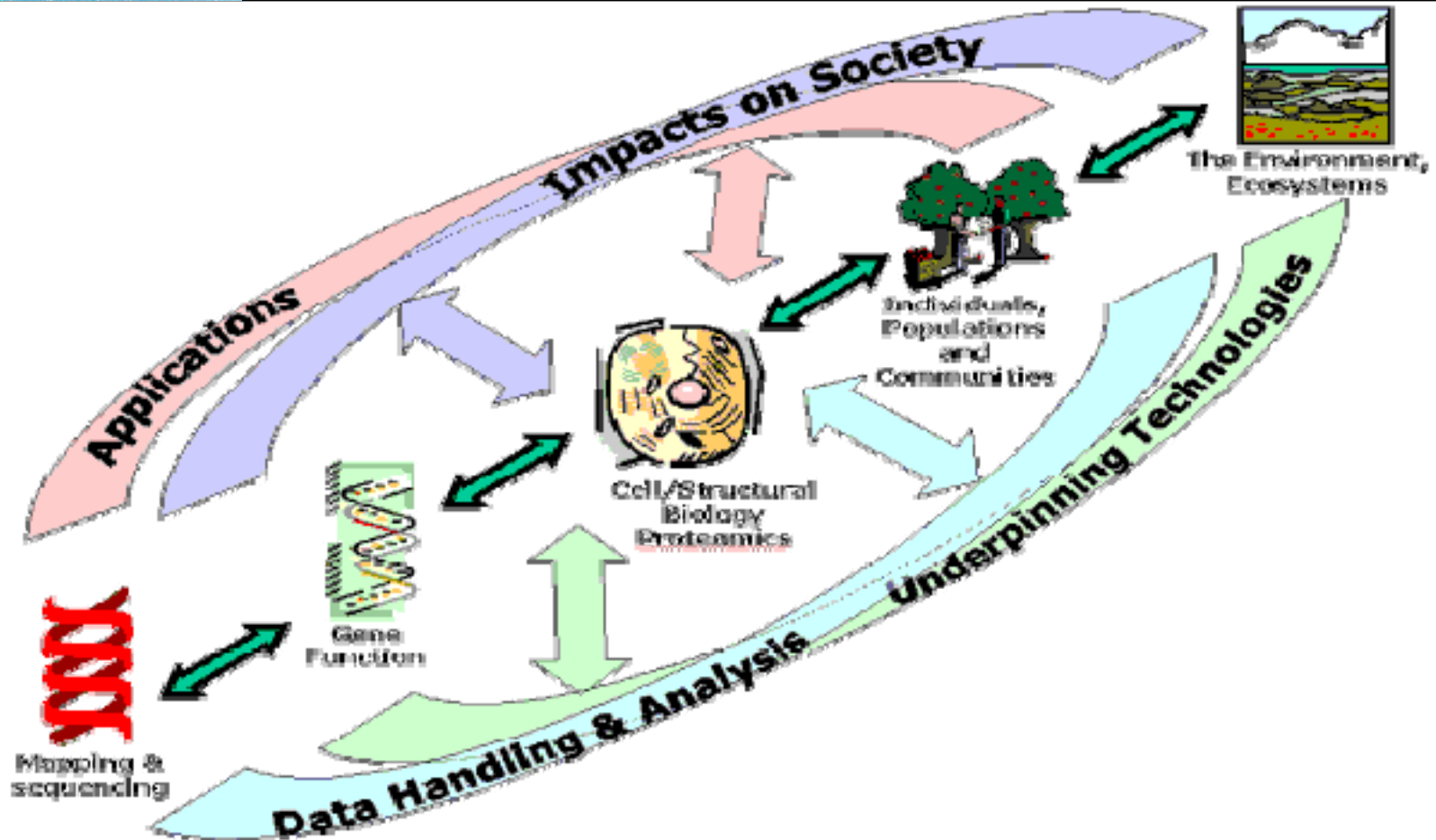
BBSRC

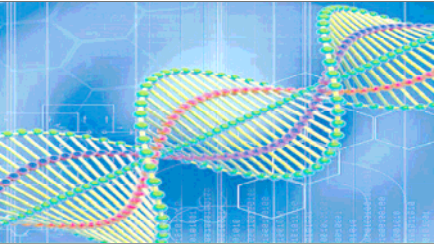
MRC

EPSRC

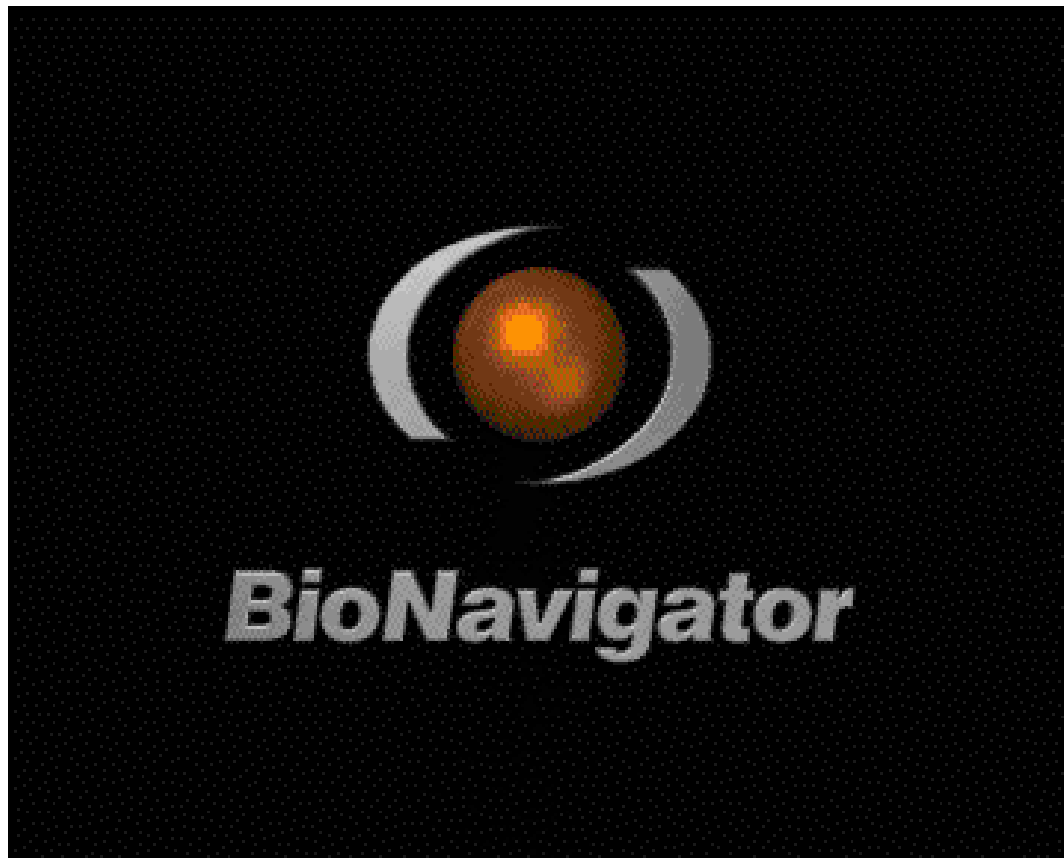
ESRC

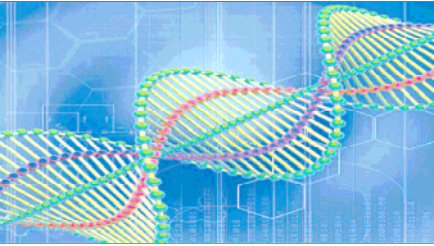
# RCUK Post-Genomic Vision





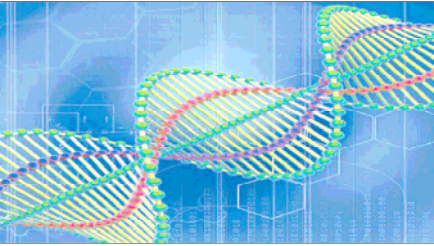
# Earth-System Science





# Scientific Objectives

- Identify ecologically important genes and traits
- Determine the ecological significance of molecular variation
  - natural variation (spatial/ temporal)
  - consequence of mans activities (pollution/ climate change)
- Determine the impact the environment has on phenotype
  - basis and theory of adaptive evolution
- Genome architecture
  - quantitative traits



# Scientific Highlights

- **Environmental Health & Pollution**
  - impact of chemicals on vertebrates and invertebrates
  - new biomarkers for exposure and adverse effect
- **Environmental Health & Disease**
  - insect responses to infection and disease
  - plant responses to infection
  - discovery of genes that slow the ageing process
- **Genes & Ecosystem services**
  - genes involved in pollutant degradation & N cycle
- **Biodiversity**
  - Genetic variation between and within species