METAFUNCTIONS:

Environmental- and meta-genomics – a bioinformatics system to detect and assign functions to habitat-specific gene patterns

Scientists are becoming increasingly interested in metagenomes – the collection of genes from all microorganisms living in a particular environment. The METAFUNCTIONS project, funded within the European Commission's FP6 NEST Adventure programme, is pooling expertise in bioinformatics, computer science, geographical information systems and marine sciences to develop a data-mining system that correlates genetic patterns in these metagenomes with contextual environmental data. This innovative tool will enable scientists to infer functions and activities for sequenced hypothetical genes, thus providing a wealth of information on niche adaptations as well as on substances with potential medical and industrial use.

Mapping out the function of genes

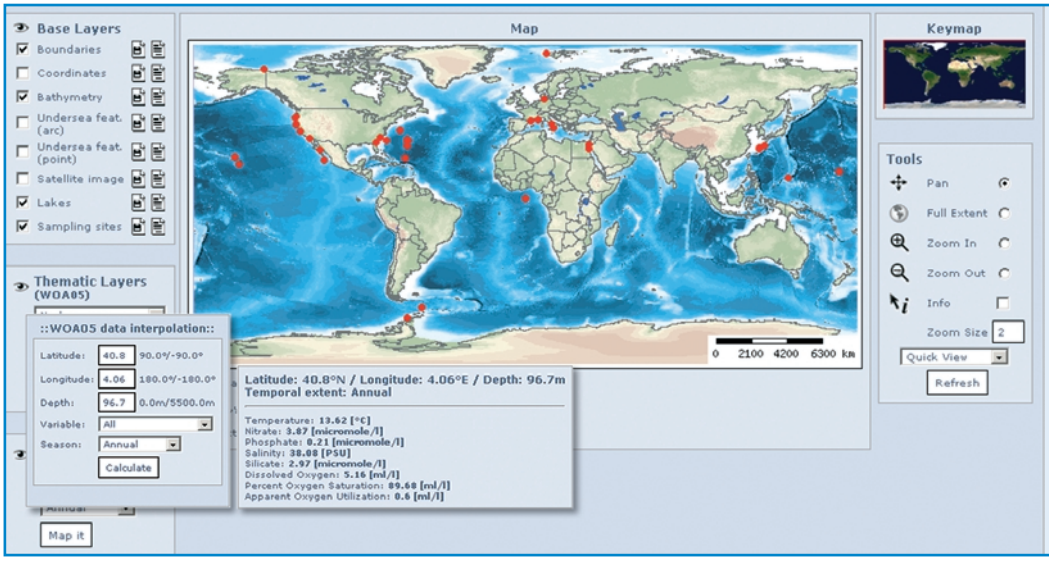
DNA sequencing has recently become a standard procedure, and with the development of high-throughput technologies, scientists can quickly sequence an organism's DNA in its entirety, esp. when it comes to microorganisms. To date (Nov. 2006) , more than 400 microbial genomes have been successfully sequenced. 'Environmentally important' marine organisms, e.g. those involved in methane production and consumption, are receiving more and more attention. Generally speaking it is difficult to culture ecologically relevant bacteria under laboratory conditions. Therefore microbial diversity and function is still largely unknown and environmental scientists have started to take DNA samples directly from the environment. The resulting sequences are known as metagenomes and they represent the genetic make-up of the microbial fraction of a particular environment.

The project will lead to innovative software tools which will reveal new correlations between functions of enzymes and proteins and marine environmental parameters.

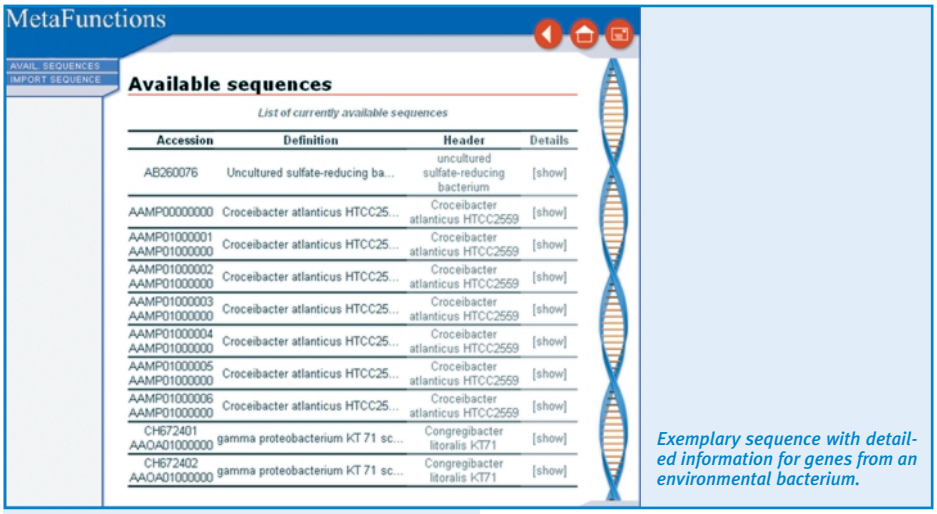
Turning data into knowledge

Recent large data acquisition projects have provided a wealth of metagenome data – however, the tools to analyse it are seriously lacking. Consequently, METAFUNCTIONS is developing a novel data-mining system to identify relationships between sequence genes and their ecological context. The ultimate aim is to help determining functions of those genes for which the activity is not yet known.

Currently, contextual data from the scientific literature as well as from global profile- and gridded data are added to our



Screenshot of the Genomes Mapserver with base and thematic layers shown on the left side, advanced navigational tools on the right side.



Exemplary sequence with detailed information for genes from an environmental bacterium.

AT A GLANCE

Official title

Environmental- and meta-genomics – a bioinformatics system to detect and assign functions to habitat-specific gene patterns

Coordinator

Germany: Max Planck Institute for Marine Microbiology

Partners

- *International: United Nations Environment Programme, Global Resource Information Database – Europe*
- *Poland: Institute of Computing Science, Poznan University of Technology*
- *Germany: Technology Transfer Centre Bremerhaven*

Further information

V.i.S.d.P. Prof. Frank Oliver Glöckner
Max Planck Institute for Marine Microbiology
Microbial Genomics Group
Celsiusstrasse 1
28359 Bremen, Germany
Tel: +49 (0)421 2028 970
Fax: +49 (0)421 2028 580
E-mail: fog@mpi-bremen.de

Duration

36 months

Project reference

Contract No. 511784 (NEST)
Web: http://www.metafunctions.org

This research project has received EU funding from the Community's Sixth Framework Programme. *This factsheet reflects only the author's views; the Community is not liable for any use that may be made of this information.*

SIXTH FRAMEWORK PROGRAMME

newly developed database hosting all available marine genome and metagenome sequences. Visualisation and access to the data is provided by our Genomes Mapserver.

Our project partners are now working on innovative software tools for smart data analysis and interpretation of the integrated data. Hundred thousands of scientific publications are screened for relevant information. A variety of natural language processing algorithms are being tested – particularly those that extract relevant information from texts – to collate this data and convert it into a structured database format. METAFUNCTIONS is also developing data-mining techniques to identify novel or interesting patterns in genomic data sets. Gene patterns – for example, the physical clustering of genes within a genome – will soon be correlated to the contextual habitat data.

The publicly available website at www.megx.net/gms allows scientists all around the world to access integrated genomic and ecological data and clearly visualise the results of their analysis.

The most innovative aspect is the inclusion of a geographic information system (GIS) to visualise and analyse data from a geographical or spatial perspective. The Genomes Mapserver, already allows the project team to access integrated genomic and ecological data. Analytical tools like MetaLook are under development to enable the project team and later scientists around the world to clearly visualise the results of diverse analyses.

Drawing on expertise

The METAFUNCTIONS approach is only possible through the integration of a diverse range of scientific disciplines. Four European and international institutions from Germany, Switzerland (UNEP) and Poland are combining their expertise in bioinformatics, computer sciences, geography, molecular and microbial ecology, and marine sciences. This innovative combination is producing new innovative software tools with broad application and high potential pay-off. In particular, this Adventure project should help to break through the current backlog in assigning function to the vast number of hypothetical genes that high-throughput genomic sequencing has produced. Finally, METAFUNCTIONS is assisting in the identification of enzymes and proteins with new and exciting activities. Marine ecology, biotechnology, medicine and many industrial sectors can all benefit from the knowledge derived from genome and metagenome data using the Genomes Mapserver.

A first visualisation how to study the habitat-specificity of genes with a "Blast against environmental containers". The seven environmental containers represent habitats and their characteristics, into which genes of interest from metagenomic data collections can be grouped according to their habitat specificity. In the example shown the DNA photolyase gene can be found in all but two of the containers.