



Uncovering the secrets of the deep

Hinxton, 30 July 2007 – The UniProt Consortium, which includes the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI), has added a new database repository for metagenomic and environmental data to its family of protein sequence databases. Metagenomics is the large-scale genomic analysis of microbes recovered from environmental samples as opposed to laboratory-grown organisms which represent only a small proportion of the microbial world. The UniProt Metagenomic and Environmental Sequences (UniMES) database currently contains the data from the Global Ocean Sampling Expedition (GOS), which was originally submitted to the International Nucleotide Sequence Databases (INSDC). The initial GOS dataset is composed of 28 million DNA sequences from oceanic microbes and predicts nearly 6 million proteins. By combining the predicted protein

sequences with automatic classification by InterPro, the EBI's integrated resource for protein families, domains and functional sites, UniMES uniquely provides free access to the array of genomic information gathered from the sampling expeditions, enhanced by links to further analytical resources.

Genomics holds the key to understanding the world around us and the metagenomic and environmental data represents a step forward in further charting genomic diversity. Rolf Apweiler from EMBL-EBI and one of the leaders of the UniProt Consortium said, "Throughout the ages, biological events have been the basis and source of medical therapies and industrially important processes. Analysing the genomes of diverse and novel species continues this adaptation of biological innovation for beneficial application". ●

Contact:

Louisa Wright PhD, EMBL-EBI Scientific Outreach Officer, Hinxton, UK, Tel: +44 1223 494665, www.ebi.ac.uk, louisa@ebi.ac.uk
Anna-Lynn Wegener, EMBL Press Officer, Heidelberg, Germany, Tel: +49 6221 387 452, www.embl.org, wegener@embl.de

About EMBL:

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 19 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom). Research at EMBL is conducted by approximately 80 independent groups covering the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, and Outstations in Hinxton (the European Bioinformatics Institute), Grenoble, Hamburg, and Monterotondo near Rome. The cornerstones of EMBL's mission are: to perform basic research in molecular biology; to train scientists, students and visitors at all levels; to offer vital services to scientists in the member states; to develop new instruments and methods in the life sciences and to actively engage in technology transfer activities. EMBL's International PhD Programme has a student body of about 170. The Laboratory also sponsors an active Science and Society programme. Visitors from the press and public are welcome.

About EBI:

The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL) and is located on the Wellcome Trust Genome Campus in Hinxton near Cambridge (UK). The EBI grew out of EMBL's pioneering work in providing public biological databases to the research community. It hosts some of the world's most important collections of biological data, including DNA sequences (EMBL-Bank), protein sequences (UniProt), animal genomes (Ensembl), three-dimensional structures (the Macromolecular Structure Database), data from microarray experiments (ArrayExpress), protein-protein interactions (IntAct) and pathway information (Reactome). The EBI hosts several research groups and its scientists continually develop new tools for the bio-computing community.

About UniMES:

The UniMES Metagenomic and Environmental Sequences database can be downloaded in FASTA format and with UniProt matches to InterPro methods from: ftp://ftp.ebi.ac.uk/pub/databases/uniprot/current_release/unimes/

About the UniProt Consortium:

The UniProt Consortium comprises the EBI, the Swiss Institute of Bioinformatics (SIB) and the Protein Information Resource (PIR). EBI, located at the Wellcome Trust Genome Campus in Hinxton, UK, hosts a large resource of bioinformatics databases and services. SIB, located in Geneva, Switzerland, is the founding centre of Swiss-Prot and maintains the ExPASy (Expert Protein Analysis System) servers that are a central resource for proteomics tools and databases. PIR, hosted by the Georgetown University Medical Center (GUMC) in Washington, DC, USA, is heir to the oldest protein sequence database, Margaret Dayhoff's Atlas of Protein Sequence and Structure published by the National Biomedical Research Foundation (NBRF) from 1965-1978. In 2002, EBI, SIB and PIR joined forces as the UniProt Consortium. This unification was made possible by funding from the U.S. National Institutes of Health, totalling US \$ 15 million over 3 years. The primary mission of the consortium is to support biological research by maintaining a high quality database that serves as a stable, comprehensive, fully classified, richly and accurately annotated protein sequence knowledgebase, with extensive cross-references and querying interfaces freely accessible to the scientific community. UniProt will build upon the solid foundations laid by the consortium members over many years.

Policy regarding use

EMBL press releases may be freely reprinted and distributed via print and electronic media. Text, photographs & graphics are copyrighted by EMBL. They may be freely reprinted and distributed in conjunction with this news story, provided that proper attribution to authors, photographers and designers is made. High-resolution copies of the images can be downloaded from the EMBL web site: www.embl.org