Press release - 9 May 2008

Issued by the Centre for Ecology & Hydrology, UK

International Group of scientists publishes guidelines for the improved description of Genome and Metagenomes

The Genomic Standards Consortium (GSC), an international initiative including representatives from a range of major sequencing and bioinformatics centres, has created a new guideline for describing genomes and metagenomes. The guideline, 'Minimum Information about a Genome Sequence' (MIGS), is published in the May issue of the journal *Nature Biotechnology*.

The new guideline is one of a number of emerging minimum information (MI) standards. Increased use of ultra-high-throughput sequencing technologies has led to the number and pace of genomic and metagenomic sequencing projects growing rapidly. Common standards such as MIGS are therefore increasingly vital to scientific progress, as groups from around the world look to share their data.

The GSC started this project to remedy the lack of descriptive information currently attached to genome and metagenome sequences in public databases. This is particularly true for environmental samples, which are amassing at an astounding pace.

According to lead author and founding member of the GSC, Dr. Dawn Field, of the UK Centre for Ecology & Hydrology, "MIGS is relatively easy to complete compared to other MI standards that are emerging. There is great enthusiasm in the community for this project and we are already collecting MIGS-compliant reports. We are a highly collaborative group and open to new participants joining the GSC at any time".

-ends-

Editors notes

Contacts:
Lead author, Dawn Field, Centre for Ecology & Hydrology – Tel: 01865 281630, email: dfield@ceh.ac.uk
Press Office, Centre for Ecology & Hydrology – Barnaby Smith – Tel: 07920 295384 Email: cehpress@ceh.ac.uk

The full paper reference is Field, D. et al. 2008. Towards a richer description of our complete collection of genomes and metagenomes: the "Minimum

Information about a Genome Sequence" (MIGS) specification. Nat. Biotechnol., May 2008**.**

The 2006 impact factor for *Nature Biotechnology* is 22.7, according to the ISI Journal Citation Reports. This places *Nature Biotechnology* first among primary research journals in the field of biotechnology.
http://www.nature.com/nbt/index.html

The Genomic Standards Consortium is an initiative working towards richer descriptions of our collection of genomes and metagenomes. Established in September 2005, this international community includes representatives from a range of major sequencing and bioinformatics centres (including NCBI, EMBL, DDBJ, JCVI, JGI, EBI, Sanger, FIG) and research institutions.

In biology the genome of an organism is its whole hereditary information and is encoded in the DNA (or, for some viruses, RNA).

Metagenomics (also Environmental Genomics, Ecogenomics or Community Genomics) is the study of genetic material recovered directly from environmental samples.

More information on the Genomic Standards Consortium can be found on the following websites:
http://gensc.org
http://darwin.nerc-oxford.ac.uk/gc_wiki/index.php/GSC
http://en.wikipedia.org/wiki/Genomic_Standards_Consortium

CEH work on Environmental Genomics:
http://www.ceh.ac.uk/sci_programmes/EnvGenomics.html

The NERC Environmental Bioinformatics Centre:
http://nebc.nox.ac.uk/

Details of MIGS

MIGS describes the genomes of eukaryotes, bacteria, archaea, plasmids, organelles, viruses and the communities of genomes found in metagenomic samples. The GSC expects MIGS to be stable enough to be used now but is open to further modification, for example as sequencing technologies change.

The MIGS specification enables description of the range of possible genomes and metagenomes using core descriptors such as origin of the nucleic acid, its source environment, date and time of sampling, and sample and sequence processing.

Specifically, the MIGS guideline describes each project as an Investigation split into a study and an assay. Under study, information can be reported about the environment of the sample (i.e. geographic location and habitat) and the details of the origin of the nucleic acid sequence. The assay section of the guideline contains a description of the use of a single sequencing technology, or a combination of technologies.

The GSC is currently working on an implementation of the MIGS specification in the form of the ‘Genomic Contextual Data Markup Language’ (GCDML). In contrast to minimum reporting, this expanded set of descriptors should facilitate ‘maximal’ reporting of information. GCDML has already been adopted by the Gordon and Betty Moore Foundation.

The Genomic Standards Consortium (GSC)

Established in September 2005, the GSC is an effort involving a wide range of participating repositories, sequencing centres, bioinformatics centres and databases. These include the NCBI, EMBL, DDBJ, JCVI, JGI, EBI, Sanger, CAMERA and a range of research institutions. The goal of GSC is to promote mechanisms for standardising the description of (meta)genomes, including the exchange and integration of (meta)genomic data.

All of these computational researchers want to improve the richness of reporting for the purpose of enabling higher quality computational analyses. This includes the ability to build better interfaces to (meta)genomic databases that offer rich options for sorting, searching, filtering, and processing groups of (meta) genomes.

Further information about the GSC and its ongoing activities can be found at: http://gensc.org.

The NIEeS and the European Bioinformatics Institute (EBI) have hosted recent GSC workshops. The UK NERC (www.nerc.ac.uk) provided funds for coordination and infrastructure building activities.