

---

## EBI becomes new host of Europe's raw sequence information

European Bioinformatics Institute receives boost to secure archives of sequencing data

**Hinxton, 14 July 2008** - Today the European Molecular Biology Laboratory's European Bioinformatics Institute (EMBL-EBI) announced that it will take guardianship of Europe's archive of raw DNA sequence data, after receiving funding from the Wellcome Trust. Currently containing 100TB of data, the Trace Archive is one of the world's largest databases of biological information. The new Archive at EMBL-EBI will be capable of accommodating the vastly increased data volume from new technology sequencing machines. The existing archive was originally created by the Wellcome Trust Sanger Institute who will continue to manage the data during the transition. The EMBL-EBI will maintain this important archive of DNA sequence in collaboration with the National Center for Biotechnology Information (NCBI) in the United States.

The seemingly simple code of As, Ts, Gs, and Cs holds secrets we are only beginning to understand. As researchers make progress in relating DNA sequence to cause and effect, we can understand life at the genomic level. For this reason, safekeeping the raw output of the sequencing machines, known as trace reads, is vital in allowing reanalysis of DNA sequence for purposes which might have been unimaginable when the sequence was first obtained. In particular, as more is known about the genetic basis of disease, existing sequence information can be reanalysed rather than samples resequenced. In a cost comparison, reanalysis can be up to 1000-fold cheaper than regenerating sequence.

The Trace Archive is a collaborative effort between Ewan Birney and Paul Flicek at the EMBL-EBI. Flicek, head of the Vertebrate Genomics Team said "The Trace Archive has been excellently maintained for the scientific community by the Wellcome Trust Sanger Institute and the EBI will perpetuate this. Our first task is to incorporate the novel types of data generated by new sequencing technologies, for example, short reads". He continued, "Because the Archive combines data from research institutions worldwide, it allows more meaningful analysis of the information it contains. This can be applied to discover patterns of variation in DNA sequence that would otherwise be undetectable".

Capturing raw sequence data will be especially important for initiatives such as the 1,000 Genomes Project, which aims to create the most detailed and medically useful catalogue of human genetic variation to date. The increase in storage capacity planned over the next four years will approximately double the amount of data currently held at the Genome Campus. This reflects the explosion in the amount of sequence data generated as technology and molecular biology advances. The trace repository currently contains over 1,770 million traces from 1262 species. ●

---

**Contact:**

[contactpress@ebi.ac.uk](mailto:contactpress@ebi.ac.uk)

Louisa Wright PhD, Scientific Outreach Officer, EMBL-EBI, Hinxton, UK. Tel: +44 (0)1223 494665, [www.ebi.ac.uk](http://www.ebi.ac.uk), [louisa@ebi.ac.uk](mailto:louisa@ebi.ac.uk)

### Trace Archive funding:

The Wellcome Trust has awarded up to €6.3 million (£5 million) to Ewan Birney and Paul Flicek from the European Bioinformatics Institute to securing the funding of the Trace Archive until 2012. The Trace Archive is accessible from <http://trace.ensembl.org/>

### 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 biocomputing community. [www.ebi.ac.uk](http://www.ebi.ac.uk)

### About EMBL:

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 20 member states (Austria, Belgium, Croatia, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the United Kingdom) and associate member state Australia. 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. [www.embl.org](http://www.embl.org)

---

### Contact:

[contactpress@ebi.ac.uk](mailto:contactpress@ebi.ac.uk)

Louisa Wright PhD, Scientific Outreach Officer, EMBL-EBI, Hinxton, UK. Tel: +44 (0)1223 494665, [www.ebi.ac.uk](http://www.ebi.ac.uk), [louisa@ebi.ac.uk](mailto:louisa@ebi.ac.uk)