

# CS Seminar

## ***European Nucleotide Archive: data, services and compression technology***

**Dr. Guy Cochrane**  
**European Nucleotide Archive**

**Date:**

February 22, 2012

Wednesday

3:00 pm

**Venue:**

Room 328

Chow Yei Ching Building

The University of Hong Kong

**Abstract:**

Aggressive technological development in nucleic acid sequencing platforms continues to bring unprecedented growth in data volumes and a broadening of the range of biological applications to which sequencing can be put. The European Nucleotide Archive (ENA; <http://www.ebi.ac.uk/ena/>) brings together a number of pre-existing and new nucleic acid sequencing-related primary data repositories, including the European node of the global Sequence Read Archive, in order to provide a comprehensive public domain data resource with global coverage and reach.

Amongst ENA responses to growth in data volumes has been the development of the reference-based sequence data compression technology, CRAM ([http://www.ebi.ac.uk/ena/about/cram\\_toolkit](http://www.ebi.ac.uk/ena/about/cram_toolkit)). In lossless mode, our technology achieves significant compression over current industry standard formats. Using a range of models for the judicious and selective reduction of quality information, we see a greater than 20-fold compression over standard formats.

In the talk, I will outline the nature of the ENA need for sequence data compression, explore the approach taken in CRAM, describe the development of models for quality information reduction and comment on the broader utility of the technology.

**About the Speaker:**

Dr. Guy Cochrane is the head of the European Nucleotide Archive (ENA; <http://www.ebi.ac.uk/ena/>). The ENA provides a comprehensive repository for public nucleotide sequence data, attracting users from a multitude of research disciplines and serving as underlying data infrastructure for numerous bioinformatics services. Under Dr. Cochrane, the team has launched important new services, such as the Sequence Read Archive for raw data from next generation sequencing platforms and the CRAMtools sequence data compression platform ([http://www.ebi.ac.uk/ena/about/cram\\_toolkit](http://www.ebi.ac.uk/ena/about/cram_toolkit)). In addition to the management of the twenty-strong team of biological curators, bioinformaticians and software engineers, he contributes editorial work to a number of journals and meetings and has been involved in standardisation activities in many areas of bioinformatics. He received his PhD in cancer cell and molecular biology from the University of East Anglia in 1999 and carried out post-doctoral work in the molecular biology of photoreception at Cambridge University prior to joining EBI.

**All are welcome!**

**For enquiries, please call 2859 2180 or email [enquiry@cs.hku.hk](mailto:enquiry@cs.hku.hk)**

**Department of Computer Science**

**The University of Hong Kong**

