Enabling easy access to DNA sequence information

May 10 2010

The European Nucleotide Archive (ENA) is launched today, consolidating three major sequence resources to become Europe's primary access point to globally comprehensive DNA and RNA sequence information. The ENA is freely available from the European Bioinformatics Institute (EMBL-EBI), a part of European Molecular Biology Laboratory.

Faster and cheaper DNA sequencing has led to previously unimaginable amounts of data being deposited in the public nucleotide sequence databases: today, ENA holds over 20 terabases of nucleotide sequence which, combined with associated information (annotation), occupies 230 terabytes of disk space. Carefully annotated and crosslinked sequence records from the EMBL Nucleotide Sequence Database (EMBL-Bank) form the backbone of the ENA. But importantly, ENA now also provides direct access to raw sequence data: the European Trace Archive contains raw data from electrophoresis-based sequencing machines and was previously maintained at the Wellcome Trust Sanger Institute; the Sequence Read Archive (SRA) is a newly established repository for raw data from next-generation (array-based) sequencing platforms. Improved submission and data-access tools make it easier for ENA's users to share their sequence data.

"Large-scale DNA sequencing was previously the domain of a small number of specialist labs, but next-generation sequencing has made it accessible to the majority of molecular life scientists," explains Graham Cameron, the EMBL-EBI's Associate Director. "The launch of ENA

reflects our continuing commitment to promoting scientific progress by providing global access to nucleotide sequence information. This has been central to EMBL's mission since the 1980s when we launched the EMBL Data Library."

Guy Cochrane, who leads the ENA team, stated that "ENA has been designed to provide our users with improved access both to annotated and to raw sequence data through the same user-friendly interface. It provides graphical browsing, web services, text search and a new rapid sequence similarity search. ENA also provides access to related information, with over 190 million cross references to external records, many of which are in other EMBL-EBI data resources."

The ENA team plans to launch many new features for the resource over the next twelve months, including enhancements to the user-friendly browser, improved interactive submissions tools and organism- and project-centred portals into ENA data.

Tim Hubbard, Head of Informatics at the Wellcome Trust Sanger Institute, said: "As major generators of DNA sequence data, it is important to us that the research community has ready access not only to annotated sequence information, but also to raw data. It's great to see the launch of ENA with new interfaces for users to this vast and rapidly growing body of information." Funding for the ENA is provided by EMBL, the Wellcome Trust and SLING, a Framework Programme 7 project coordinated by the EMBL-EBI and funded by the European Commission.

Provided by European Molecular Biology Laboratory