

Public collections of DNA and RNA sequence reach 100 gigabases

August 22 2005

The world's three leading public repositories for DNA and RNA sequence information have reached 100 gigabases [100,000,000,000 bases; the 'letters' of the genetic code] of sequence.

Thanks to their data exchange policy, which has paved the way for the global exchange of many types of biological information, the three members of the International Nucleotide Sequence Database Collaboration [INSDC, www.insdc.org] – EMBL Bank [Hinxton, UK], GenBank [Bethesda, USA] and the DNA Data Bank of Japan [Mishima, Japan] all reached this milestone together.

Graham Cameron, Associate Director of EMBL's European Bioinformatics Institute, says "This is an important milestone in the history of the nucleotide sequence databases. From the first EMBL Data Library entry made available in 1982 to today's provision of over 55 million sequence entries from at least 200,000 different organisms, these resources have anticipated the needs of molecular biologists and addressed them – often in the face of a serious lack of resources."

David Lipman, Director of the National Center for Biotechnology Information, adds: "Today's nucleotide sequence databases allow researchers to share completed genomes, the genetic make-up of entire ecosystems, and sequences associated with patents. The INSDC has realized the vision of the researchers who initiated the sequence database projects, by making the global sharing of nucleotide sequence information possible."

Takashi Gojobori, Director of the Center for Information Biology and DNA Data Bank of Japan, says: "The INSDC has laid the foundations for the exchange of many types of biological information. As we enter the era of systems biology and researchers begin to exchange complex types of information, such as the results of experiments that measure the activities of thousands of genes, or computational models of entire processes, it is important to celebrate the achievements of the three databases that pioneered the open exchange of biological information."

In the late 1970s, as researchers started to study organisms at the level of their genetic code, several groups began to explore the possibility of developing a public repository for sequence information. In the early 1980s this led to the launch of two databases: the first was the EMBL Data Library, based at the European Molecular Biology Laboratory [EMBL] in Heidelberg, Germany [the Data Library is now known as EMBL Bank and is based at EMBL's European Bioinformatics Institute, Hinxton, UK]. Hot on its heels came GenBank, initially hosted by the Los Alamos National Laboratory [LANL] and now based at the National Center for Biotechnology Information, Bethesda, MD, USA. Both of these databases were seeded by collections begun by far-sighted individuals: EMBL Bank by the collection of Kurt Stüber, then based at the University of Cologne in Germany, and GenBank by the collection of Walter Goad at LANL.

The two nascent databases began collaborating very early on, an interaction that was initiated by Greg Hamm, the EMBL Data Library's first employee. Staff at the two databases, which at that time had to find sequences in published journal articles and re-key them into the databases, allocated journals to each team to avoid duplication of effort, and began the arduous task of mapping the fields from one database onto those of the other so that they could exchange information. By the time the International Nucleotide Sequence Consortium became formalized in February 1987, a third partner, the DNA Data Bank of Japan, had been

launched at the National Institute of Genetics in Mishima, and collaborated with its European and US counterparts right from the start.

Much has changed since the days when sequences were manually keyed in from the literature or sent on floppy disc and distributed to users on 9-track magnetic tapes, but the purpose of the databases – to make every nucleotide sequence in the public domain freely available to the scientific community as rapidly as possible – remains as strong now as it was in the beginning.

Source: European Molecular Biology Laboratory (EMBL)

Citation: Public collections of DNA and RNA sequence reach 100 gigabases (2005, August 22) retrieved 26 April 2024 from <https://phys.org/news/2005-08-dna-rna-sequence-gigabases.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.