

EMBL-EBI researchers make DNA storage a reality

Every film and TV programme ever created - in a teacup

Hinxton, 23 January 2013 - Researchers at the EMBL-European Bioinformatics Institute (EMBL-EBI) have created a way to store data in the form of DNA – a material that lasts for tens of thousands of years. The new method, published today in the journal *Nature*, makes it possible to store at least 100 million hours of high-definition video in about a cup of DNA. There is a lot of digital information in the world – about three zettabytes' worth (that's 3000 billion billion bytes) – and the constant influx of new digital content poses a real challenge for archivists. Hard disks are expensive and require a constant supply of electricity, while even the best 'no-power' archiving materials such as magnetic tape degrade within a decade. This is a growing problem in the life sciences, where massive volumes of data – including DNA sequences – make up the fabric of the scientific record.

"We already know that DNA is a robust way to store information because we can extract it from bones of woolly mammoths, which date back tens of thousands of years, and make sense of it," explains Nick Goldman of EMBL-EBI. "It's also incredibly small, dense and does not need any power for storage, so shipping and keeping it is easy."

Reading DNA is fairly straightforward, but writing it has until now been a major hurdle to making DNA storage a reality. There are two challenges: first, using current methods it is only possible to manufacture DNA in short strings. Secondly, both writing and reading DNA are prone to errors, particularly when the same DNA letter is repeated. Nick Goldman and co-author Ewan Birney, Associate Director of EMBL-EBI, set out to create a code that overcomes both problems.

"We knew we needed to make a code using only short strings of DNA, and to do it in such a way that creating a run of the same letter would be impossible. So we figured, let's break up the code into lots of overlapping fragments going in both directions, with indexing information showing where each fragment belongs in the overall code, and make a coding scheme that doesn't allow repeats. That way, you would have to have the same error on four different fragments for it to fail – and that would be very rare," says Ewan Birney.

The new method requires synthesising DNA from the encoded information: enter Agilent Technologies, Inc, a California-based company that volunteered its services. Ewan Birney and Nick Goldman sent them encoded versions of: an .mp3



Nick Goldman of EMBL-EBI, looking at synthesised DNA.
Credit: EMBL Photolab

of Martin Luther King's speech, "I Have a Dream"; a .jpg photo of EMBL-EBI; a .pdf of Watson and Crick's seminal paper, "Molecular structure of nucleic acids"; a .txt file of all of Shakespeare's sonnets; and a file that describes the encoding.

"We downloaded the files from the Web and used them to synthesise hundreds of thousands of pieces of DNA – the result looks like a tiny piece of dust," explains Emily Leproust of Agilent. Agilent mailed the sample to EMBL-EBI, where the researchers were able to sequence the DNA and decode the files without errors.

"We've created a code that's error tolerant using a molecular form we know will last in the right conditions for 10 000 years, or possibly longer," says Nick Goldman. "As long as someone knows what the code is, you will be able to read it back if you have a machine that can read DNA."

Although there are many practical aspects to solve, the inherent density and longevity of DNA makes it an attractive storage medium. The next step for the researchers is to perfect the coding scheme and explore practical aspects, paving the way for a commercially viable DNA storage model. ●

Source Article

Toward practical high-capacity low-maintenance storage of digital information in synthesised DNA. Advanced online publication in *Nature* on 23 January 2013; DOI: 10.1038/nature11875

Contact:

Mary Todd-Bergman, Senior Communications Officer, Hinxton, UK, Tel: +44 1223 494 665, www.ebi.ac.uk, contactpress@ebi.ac.uk
Isabelle Kling, EMBL Communication Officer, Heidelberg, Germany, Tel: +49 6221 387 8355, www.embl.org, isabelle.kling@embl.de

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 85 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. Around 190 students are enrolled in EMBL's International PhD programme. Additionally, the Laboratory offers a platform for dialogue with the general public through various science communication activities such as lecture series, visitor programmes and the dissemination of scientific achievements.

About EMBL-EBI:

The European Bioinformatics Institute is part of EMBL, Europe's flagship laboratory for the life sciences. EMBL-EBI provides freely available data from life science experiments covering the full spectrum of molecular biology. About 20% of our institute is devoted to investigator-led research using computational approaches to unravel the secrets of life. Our extensive training programme helps researchers in academia and industry to make the most of the incredible amount of data being produced every day in life science experiments. We are a non-profit, intergovernmental organisation funded by EMBL member states. Our 520 staff hail from 43 countries, and we welcome a regular stream of visiting scientists throughout the year. We are located on the Wellcome Trust Genome Campus in Hinxton, Cambridge in the United Kingdom.

Policy regarding use

EMBL press and picture releases including photographs, graphics, movies and videos are copyrighted by EMBL. They may be freely reprinted and distributed for non-commercial use via print, broadcast and electronic media, provided that proper attribution to authors, photographers and designers is made. High-resolution copies of the images can be downloaded from the EMBL web site: www.embl.org.