

Researchers make DNA storage a reality

January 23 2013



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

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 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.

More information: Towards practical, high-capacity, low-maintenance information storage in synthesized DNA - Nick Goldman, Paul Bertone, Siyuan Chen, Christophe Dessimoz, Emily M. LeProust, Botond Sipos & Ewan Birney - Advanced online publication in *Nature* on 23 January, 2013. [DOI:10.1038/nature11875](https://doi.org/10.1038/nature11875)

Abstract

Digital production, transmission and storage have revolutionized how we access and use information but have also made archiving an increasingly complex task that requires active, continuing maintenance of digital media. This challenge has focused some interest on DNA as an attractive target for information storage because of its capacity for high-density information encoding, longevity under easily achieved conditions and proven track record as an information bearer. Previous DNA-based information storage approaches have encoded only trivial amounts of

information or were not amenable to scaling-up, and used no robust error-correction and lacked examination of their cost-efficiency for large-scale information archival. Here we describe a scalable method that can reliably store more information than has been handled before. We encoded computer files totalling 739 kilobytes of hard-disk storage and with an estimated Shannon information of 5.23106 bits into a DNA code, synthesized this DNA, sequenced it and reconstructed the original files with 100% accuracy. Theoretical analysis indicates that our DNA-based storage scheme could be scaled far beyond current global information volumes and offers a realistic technology for large-scale, long-term and infrequently accessed digital archiving. In fact, current trends in technological advances are reducing DNA synthesis costs at a pace that should make our scheme cost-effective for sub-50-year archiving within a decade.

Provided by European Molecular Biology Laboratory

Citation: Researchers make DNA storage a reality (2013, January 23) retrieved 19 April 2024 from <https://phys.org/news/2013-01-dna-storage-reality.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.
