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

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.

Journal reference: Nature

Provided by European Molecular Biology Laboratory

view popular send feedback to editors
5 /5 (3 votes)

- 1
- 2
- 3
- 4
- 5

Move the slider to adjust rank threshold, so that you can hide some of the comments.

Display comments: newest first

jalmy
1 hour ago

Rank: not rated yet
Neat.

  • report

NeutronicallyRepulsive
1 hour ago

Rank: not rated yet
Well, what if part of someone's DNA is "The Lord Of The Rings.mp4" will the owner be sued by
copyright violation on multiple accounts (times cells in a body)? Also can we keep the self replicating
part, put 2001 - Space Odyssey, and Alien movies as an appendage DNA, and let the organisms mate with
each other while Rotten Tomatoes rating will be the fitness function? Suddenly, the future looks much brighter.

- report

**Antoweif**
37 minutes ago

Rank: **not rated yet**
Nice. What are the cons?

- report

Please register or sign in to add a comment. Registration is free, and takes less than a minute. Read more

Sign in with

Email

Password

Sign In

Forgot your password? Click here to reset it.
Notify me via email of follow-up comments posted here Sign in first

Rank

- 1
- 2
- 3
- 4
- 5

5 / 5 (3 votes)

- Featured
- Last comments
- Popular
- Most shared
- Partners

- A new world record for solar cell efficiency Jan 17, 2013 | 5 / 5 (47) | 49
- 'Quadruple helix' DNA discovered in human cells Jan 20, 2013 | 4.6 / 5 (49) | 18
- Mathematical breakthrough sets out rules for more effective teleportation Jan 16, 2013 | 4.4 /
Nearly perfect, ultrathin invisibility cloak could have wide practical applications

Jan 17, 2013 | 4.5 / 5 (30) | 21

Japan to replace nuclear plant with world's largest wind farm

Jan 18, 2013 | 3.8 / 5 (35) | 99

DNA used to encode a book and other digital information

May 10, 2010 | 0

Enabling easy access to DNA sequence information

May 20, 2011 | 0

DNA falls apart when you pull it

Sep 12, 2012 | 0

Information theory helps unravel DNA's genetic code

Tags
woolly mammoths, dna sequences, digital information, digital content, magnetic tape, dna, journal nature

Relevant PhysicsForums posts

- Quadruple-Helix DNA Found in Human Cells
  Jan 20, 2013
- Quantitative PEMF therapy data please
  Jan 20, 2013
- What is a meta nucleus in context of biology?
  Jan 19, 2013
- reversing/curing hearing loss (part 2)
  Jan 18, 2013
- solar powered vertebrate (through symbiosis with algae)
  Jan 18, 2013
- Are humans animals?
  Jan 16, 2013
- More from Physics Forums - Biology

More news stories

How salt stops plant growth

Until now it has not been clear how salt, a scourge to agriculture, halts the growth of the plant-root system. A team of researchers, led by the Carnegie Institution's José Dinneny and Lina Duan, found that ...
Better outlook for dwindling black macaque population in Indonesia

Since at least the 1970s, the population of critically endangered Sulawesi black macaques living in an Indonesian nature reserve has been dropping. But a new study by researchers at the University of Washington ...

Underwater CO2 shows potential as barrier to Asian carp

As the Asian carp population grows and the threat of the invasive species entering Lake Michigan through one of the Chicago canals is monitored, a University of Illinois researcher believes using two barrier ...

Parasites of Madagascar's lemurs expanding with climate change

Rising temperatures and shifting rainfall patterns in Madagascar could fuel the spread of lemur parasites and the diseases they carry.

Chimpanzees: Hormone oxytocin likely to play key role in maintaining social relations with cooperation partners

Animals which maintain cooperative relationships show gains in longevity and offspring survival. However, little is known about the cognitive or hormonal mechanisms involved in cooperation. Researchers of ...

Blocking digestive enzymes may reverse shock, stop multiorgan failure

New research from the University of California, San Diego published in the Jan. 23 issue of Science Translational Medicine moves researchers closer to understanding and developing treatments for shock, sepsis ...

New method of producing nanomagnets for information technology

An international team of researchers has found a new method of producing molecular magnets. Their thin layer systems made of cobalt and an organic material could pave the way for more powerful storage media ...
**Personal epigenetic 'signatures' found consistent in prostate cancer patients' metastases**

In a genome-wide analysis of 13 metastatic prostate cancers, scientists at the Johns Hopkins Kimmel Cancer Center found consistent epigenetic "signatures" across all metastatic tumors in each patient. The discovery of the ...

**Researchers discover promising prognostic marker for aggressive breast cancer**

A team of researchers led by Goutham Narla, MD, PhD, at Case Western Reserve University School of Medicine and University Hospitals Case Medical Center, and collaborators at the Mount Sinai School of Medicine and Erasmus ...  

**Health and environment: A closer look at plastics**

Plastics have transformed modern society, providing attractive benefits but also befouling waterways and aquifers, depleting petroleum supplies and disrupting human health.

**New way to identify 'smoked' grapes and wines**

With climate change sparking concern about an increased risk of wildfires, scientists are reporting development of a way to detect grapes exposed to smoke from those fires, which otherwise could be vented ...