

DNA Data Storage next Quantum Leap

by Geraint Jones via digby Monday, Aug 20 2012, 10:29am
international / prose / post

Entire internet would fit on a thumb drive

The amazing data storage capability of DNA notwithstanding, this story is primarily extolling the virtues of extra-dimensional advantages. The primary reason DNA dwarfs known digital storage methods is simply due to the fact that DNA stores its information in THREE dimensions whereas today's conventional digital storage methods only encode data in two dimensional form.



From The [Guardian](#) UK:

Scientists have for the first time used DNA to encode the contents of a book. At 53,000 words, and including 11 images and a computer program, it is the largest amount of data yet stored artificially using the genetic material.

The researchers claim that the cost of DNA coding is dropping so quickly that within five to 10 years it could be cheaper to store information using this method than in conventional digital devices.

Deoxyribonucleic acid or DNA – the chemical that stores genetic instructions in almost all known organisms – has an impressive data capacity. One gram can store up to 455bn gigabytes: the contents of more than 100bn DVDs, making it the ultimate in compact storage media.

A three-strong team led by Professor George Church of Harvard Medical School has now demonstrated that the technology to store data in DNA, while still slow, is becoming more practical. They report in the journal Science that the 5.27 megabit collection of data they stored is more than 600 times bigger than the largest dataset previously encoded this way.

Writing the data to DNA took several days. "This is currently something for archival storage," explained co-author Dr Sriram Kosuri of Harvard's Wyss Institute, "but the timing is continually improving."

DNA has numerous advantages over traditional digital storage media. It can be easily copied, and is often still readable after thousands of years in non-ideal conditions. Unlike ever-changing electronic storage formats such as magnetic tape and DVDs, the fundamental techniques required to read and write DNA information are as old as life on Earth.

The researchers, who have filed a provisional patent application covering the idea, used off-the-shelf components to demonstrate their technique.

To maximise the reliability of their method, and keep costs down, they avoided the need to create very long sequences of code – something that is much more expensive than creating lots of short chunks of DNA. The data was split into fragments that could be written very reliably, and was accompanied by an address book listing where to find each code section.

Digital data is traditionally stored as binary code: ones and zeros. Although DNA offers the ability to use four "numbers": A, C, G and T, to minimise errors Church's team decided to stick with binary encoding, with A and C both indicating zero, and G and T representing one.

The sequence of the artificial DNA was built up letter by letter using existing methods with the string of As, Cs, Ts and Gs coding for the letters of the book.

The team developed a system in which an inkjet printer embeds short fragments of that artificially synthesised DNA onto a glass chip. Each DNA fragment also contains a digital address code that denotes its location within the original file.

The fragments on the chip can later be "read" using standard techniques of the sort used to decipher the sequence of ancient DNA found in archeological material. A computer can then reassemble the original file in the right order using the address codes.

The book – an HTML draft of a volume co-authored by the team leader – was written to the DNA with images embedded to demonstrate the storage medium's versatility.

DNA is such a dense storage system because it is three-dimensional. Other advanced storage media, including experimental ones such as positioning individual atoms on a surface, are essentially confined to two dimensions.

The work did not involve living organisms, which would have introduced unnecessary complications and some risks. The biological function of a cell could be affected and portions of DNA not used by the cell could be removed or mutated. "If the goal is information storage, there's no need to use a cell," said Kosuri.

The data cannot be overwritten but, given the storage capacity, that is seen as a minor issue. The exercise was not completely error-free, but of the 5.27m bits stored, only 10 were found to be incorrect. The team suggests common error-checking techniques could be implemented in future, including multiple copies of the same information so mistakes can be easily identified.

The costs of DNA-handling tools are not yet competitive enough to make this a large-scale storage medium. But the costs and scale of the tools are dropping much more quickly than their electronic equivalents. For example, handheld DNA sequencers are becoming available, which the authors suggest should greatly simplify information stored in DNA.

Kosuri foresees this revolution in DNA technologies continuing. "We may hit a wall, but there's no fundamental reason why it shouldn't continue."

© 2012 Guardian News and Media Limited

See also:

<http://tinyurl.com/8flkjsq>

