

DNA data storage breaks records

A book complete with illustrations has been encoded in DNA.

Monya Baker

16 August 2012

A trio of researchers has encoded a draft of a whole book into DNA. The 5.27-megabit tome contains 53,246 words, 11 JPG image files and a JavaScript program, making it the largest piece of non-biological data ever stored in this way.

DNA has the potential to store huge amounts of information. In theory, two bits of data can be incorporated per nucleotide — the single base unit of a DNA string — so each gram of the double-stranded molecule could store 455 exabytes of data (1 exabyte is 10^{18} bytes). Such dense packing outstrips inorganic data-storage devices such as flash memory, hard disks or even storage based on quantum-computing methods.

The book, which is fittingly a treatise on synthetic biology, was encoded by geneticists George Church and Sriram Kosuri at the Wyss Institute for Biologically Inspired Engineering in Boston, Massachusetts, and Yuan Gao, a biomedical engineer at Johns Hopkins University in Baltimore, Maryland. They report their work in *Science*¹ this week.

It marks a significant gain on previous projects — the largest of which encoded less than one-six-hundredth of the data — but organic flash drives are still many years away. There are a number of reasons why the method is not practical for everyday use. For example, both storing and retrieving information currently require several days of lab work, spent either synthesizing DNA from scratch or sequencing it to read the data.

The work illustrates the potential of nonconventional approaches, says Stuart Parkin, who is developing dense forms of inorganic storage media at the IBM-Stanford Spintronic Science and Applications Center in San Jose, California. "You could say that the physical sciences have exhausted the playground of concepts, and we now need to go beyond that world," he says. "This coupling of the biological world to the physical world will lead to some very interesting storage devices in the next decade."

Short and sweet

Encoding the DNA book didn't involve fundamentally new technology so much as the creative application of existing techniques, explains Anne Condon, a computer scientist at the University of British Columbia in Vancouver, who studies how DNA molecules can be used in



LAGUNA DESIGN/SCIENCE PHOTO LIBRARY

A double-stranded DNA molecule

computing.



Trinity Mirror / Mirrorpix / Alamy

Employees of The British Library put cleaned and repaired books back on the shelves in 1927.

Previous attempts to store information in DNA have been held up by difficulties in making perfect long strands. Shorter molecules present less of a challenge, so Church and his colleagues kept their storage strands a mere 159 nucleotides long, and generated multiple copies of each to make catching and correcting mutations easier.

In each single strand, 96 nucleotides represented the encoded data as digital ones and zeroes; 19 nucleotides showed how these data blocks should be ordered; and 44 nucleotides enabled easier sequencing. The researchers' binary code assigned 'zero' to two types of nucleotide (As and Cs) and 'one' to the other two types (Gs and Ts).

"It's using some simple ideas in very elegant ways to improve the density of information that one can store," says Condon. She says that the technology will work best for specialized applications in which data need to be stored for a long time without being read.

The ideal storage period might be as long as centuries, says Kosuri. Even as other storage technologies become as obsolete as magnetic tape and floppy disks are now, researchers will always be trying to improve technology for reading and writing DNA, because the molecule is so central to biology.

And that will bring costs down over time, enabling DNA data storage to move beyond the realm of demonstration projects. The cost of sequencing technologies has already fallen to about a thousandth of what it was four years ago, says Kosuri, and DNA synthesis has achieved the same drops over the past eight years. "The DNA chip we used for this paper held 55,000 oligonucleotides," he adds. "The newest ones hold a million."

Nature | doi:10.1038/nature.2012.11194

References

1. Church, G. M., Gao, Y. & Kosuri, S. *Science* <http://dx.doi.org/10.1126/science.1226355> (2012).