

BBC NEWS**SCIENCE & ENVIRONMENT**

23 January 2013 Last updated at 18:03

DNA 'perfect for digital storage'

By Jonathan Amos
Science correspondent, BBC News

Scientists have given another eloquent demonstration of how DNA could be used to archive digital data.

The UK team encoded a scholarly paper, a photo, Shakespeare's sonnets and a portion of Martin Luther King's I Have A Dream speech in artificially produced segments of the "life molecule".

The information was then read back out with 100% accuracy.

It is possible to store huge volumes of data in DNA for thousands of years, the researchers [write in Nature magazine](#).

They acknowledge that the costs involved in synthesizing the molecule in the lab make this type of information storage "breathtakingly expensive" at the moment, but argue that newer, faster technologies will soon make it much more affordable, especially for long-term archiving.

"One of the great properties of DNA is that you don't need any electricity to store it," explained team-member Dr Ewan Birney from the European Bioinformatics Institute (EBI) at Hinxton, near Cambridge.

"If you keep it cold, dry and dark - DNA lasts for a very long time. We know that because we routinely sequence woolly mammoth DNA that is kept by chance in those sorts of conditions." Mammoth remains are many thousands of years old.

The group cites government and historical records as examples of data that could benefit from the molecular storage option.

Much of this information is not required every day but still needs to be kept. Once encoded in DNA, it could be put away safely in a vault until it was needed.

And unlike other storage media presently in use such as hard disk-drives and magnetic tapes, the DNA "library" would not demand constant maintenance.

In addition, the universality of the life molecule means there would probably never be a backwards-compatibility issue where the technology of the day was incapable of reading the vault's archives.

"We think there will always be DNA-reading technology so long as there is DNA-based life around on Earth, assuming it is technologically sophisticated of course," Dr Birney told BBC News.

This is not the first time that DNA has been used to encode the sort of routine information we keep on our computers.

Last year, for example, an American group [published the results of a very similar experiment in Science Magazine](#). The Boston researchers laid down a whole book in DNA.

The EBI study uses slightly different techniques to achieve its goals, but has also looked deeper into some of the issues of

scalability and practicality.

Underpinning all these approaches is the exploitation of the nucleobase sequence at the heart of DNA.

The helical molecule is famously held together by four chemical groups, or nucleobases, which, when arranged in a specific order, carry the genetic instructions needed by a living organism to build and maintain itself.

The EBI storage system uses the same four "letters" but in a completely different "language" to the one understood by life.

To copy a computer file, such as a text document, the binary digits (zeros and ones) that would ordinarily represent that information on a hard drive first have to be translated into the team's bespoke code. A standard DNA synthesis machine then churns out the corresponding sequence.

But it is not one long molecule. Rather, it is multiple copies of overlapping fragments, with each fragment also carrying some indexing details that identify where in the overall sequence it should sit.

This builds redundancy into the system, meaning that if some fragments become corrupted, the data will not be lost.

Again, the same standard equipment used in molecular biology labs to read the DNA of organisms is used to pull out the information so that it can be displayed on a computer screen once more.

For its experiment, the EBI team encoded a 26-second snippet of Martin Luther King's classic anti-racism address from 1963, a ".jpg" photo of the EBI (see right); a ".pdf" of the seminal 1953 paper by Crick and Watson describing the structure of DNA, ".txt" file containing all of Shakespeare's sonnets; and a file about the encoding system itself (a total equivalent on a computer drive to about 760 kilobytes).

Physically, the DNA carrying all that information is no bigger than a speck of dust.

Team member Nick Goldman said the molecule was an incredibly dense storage medium. One gram of DNA ought to be able to hold about two petabytes of data, he added - the equivalent of about three million CDs.

Dr Goldman addressed the concern some people might have that artificial DNA code could somehow go wild and end up in the genome of another living organism.

"The DNA we've created can't be incorporated accidentally into a genome; it uses a completely different code to what the cells of living bodies use," he explained.

"And if you did end up with any of this DNA inside you, it would just be degraded and disposed of. It really has no place in a living being."

Jonathan.Amos-INTERNET@bbc.co.uk and follow me on Twitter: [@BBCAmos](https://twitter.com/BBCAmos)

[More Science & Environment stories](#)



[Dog evolved 'on the waste dump'](#)

[\[/news/science-environment-21142870\]](#)

A genetic study indicates the ability to thrive on the starchy food leftovers of early farmers was a key step in the domestication of dogs from wolves.

[Farm double payments clear hurdle](#)

[\[/news/science-environment-21171472\]](#)

[Controversial bird flu work resumes](#)

[\[/news/health-21165288\]](#)



BBC © 2013 The BBC is not responsible for the content of external sites. [Read more.](#)