

THE WALL STREET JOURNAL

WSJ.com

U.S. NEWS | Updated January 24, 2013, 8:41 a.m. ET

Storing Digital Data in DNA

Technique One Day May Replace Hard Drives as Web Leads to Information Deluge

By GAUTAM NAIK



Scientists have used DNA stored audio and text on fragments of DNA and then retrieved them with near-perfect fidelity—a technique that may provide a way to handle data in the digital age. WSJ's Gautam Naik reports.

Scientists have stored audio and text on fragments of DNA and then retrieved them with near-perfect fidelity—a technique that eventually may provide a way to handle the overwhelming data of the digital age.

The scientists encoded in DNA—the recipe of life—an audio clip of Martin Luther King Jr.'s "I Have a Dream" speech, a photograph, a copy of Francis Crick and James Watson's famous "double helix" scientific paper on DNA from 1953 and Shakespeare's 154 sonnets. They later were able to retrieve them with 99.99% accuracy.

The experiment was reported Wednesday in the journal *Nature*.

"All we're doing is adapting what nature has hit upon—a very good way of storing information," said Nick Goldman, a computational biologist at the European Bioinformatics Institute in Hinxton, England, and lead author of the *Nature* paper.

Companies, governments and universities



Getty Images

A copy of Crick and Watson's famous "double helix" scientific paper from 1953 was among items scientists successfully encoded. Above, a DNA model.

face an enormous challenge storing the ever-growing flood of digitized information, the videos, books, movies and songs sent over the Internet.

Some experts have looked for answers in biology. In recent years, they have found ways to encode trademarks in cells and poetry in bacteria, as well as store snippets of music in the genetic code of micro-organisms. But these biological things eventually die.

By contrast, DNA—the molecule that contains the genetic instructions for all living things—

is stable, durable and dense. Because DNA isn't alive, it could sit passively in a storage device for thousands of years.

Among today's data-storage devices, magnetic tapes can degrade within a decade, while hard disks are expensive and need a constant supply of electricity to hold their information, creating huge need for power for the "data farms" behind cloud computing.

DNA could hold vastly more information than the same surface volume of a disk drive—a cup of DNA theoretically could store about 100 million hours of high-definition video.

While DNA-based storage remains a long way from being commercially viable—high cost is one major hurdle—the scientific barriers are starting to fall.

Last August, researchers at Harvard University reported in the journal *Science* the encoding of an entire 54,000-word book in strands of DNA.

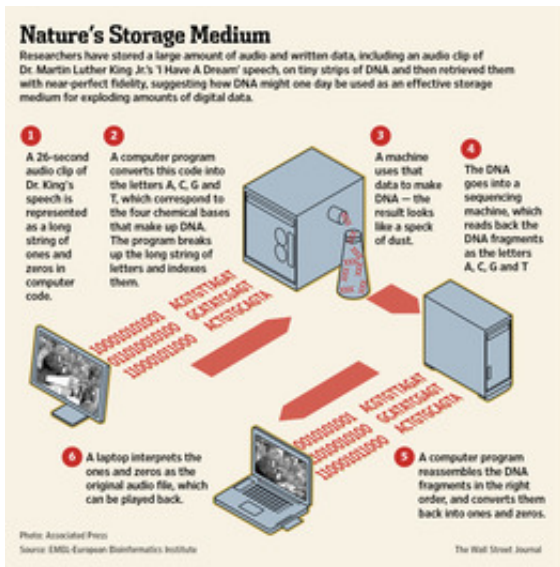
"The experiments are very similar," said George Church, a molecular geneticist at Harvard and senior researcher for the project reported in *Science*. "Because these are truly independent efforts, we've shown there's a real field here rather than just one group."

Both experiments encoded similar amounts of information and had roughly similar accuracy rates, according to Dr. Church.

The European Bioinformatics Institute is part of the European Molecular Biology Laboratory, Europe's flagship life-sciences lab. The EMBL is funded by public research money from 20 European member states.

In their experiment, Dr. Goldman and his colleagues first downloaded onto a computer a 26-second clip of Dr. King's "I Have a Dream" speech, the sonnets and the other things to be stored. The data was in normal computer code—a long string of ones and zeros.

A software program devised by Dr. Goldman's



team converted those ones and zeros into the letters A, C, G and T, the four chemical bases that make up DNA.

The single, long DNA-based string was chopped up into about 150,000 fragments, each 120 letters long. Each fragment contained about 100 letters encoding the data. The remaining 20 letters served as an index—instructions for later restoring the fragments in the right order.

The information was sent to [Agilent Technologies](#) Inc. of Santa Clara, Calif., where a laboratory machine used the data and

appropriate chemicals to manufacture physical strings of actual DNA. Those fragments were shipped to Dr. Goldman's lab in England.

"I thought the vial was empty when it arrived," said Dr. Goldman. But the DNA was there—it lay like a speck of dust at the bottom of the vial, almost impossible to see.

After some lab work, the DNA was dispatched to an EMBL lab in Heidelberg, Germany. There, a DNA-sequencing machine fired lasers at the fragments and read their genetic code, yielding a computer file in the form of As, Cs, Gs and Ts.

Back in Hinxton, a computer program reassembled the fragments in the right order, and then converted them back into ones and zeros. When run on a laptop, those ones and zeros were interpreted as the audio clip, sonnets and other items. When the clip of Dr. King's speech was played back, it sounded just like the original version, said Dr. Goldman.

Plenty of challenges remain before DNA storage could become a cheap and reliable commercial process.

"In 10 years, it's probably going to be about 100 times cheaper," said Dr. Goldman. "At that time, it probably becomes economically viable."

Write to Gautam Naik at gautam.naik@wsj.com

A version of this article appeared January 24, 2013, on page A3 in the U.S. edition of The Wall Street Journal, with the headline: Storing Digital Data in DNA.

Copyright 2012 Dow Jones & Company, Inc. All Rights Reserved

This copy is for your personal, non-commercial use only. Distribution and use of this material are governed by our [Subscriber Agreement](#) and by copyright law. For non-personal use or to order multiple copies, please contact Dow Jones Reprints at 1-800-843-0008 or visit www.djreprints.com