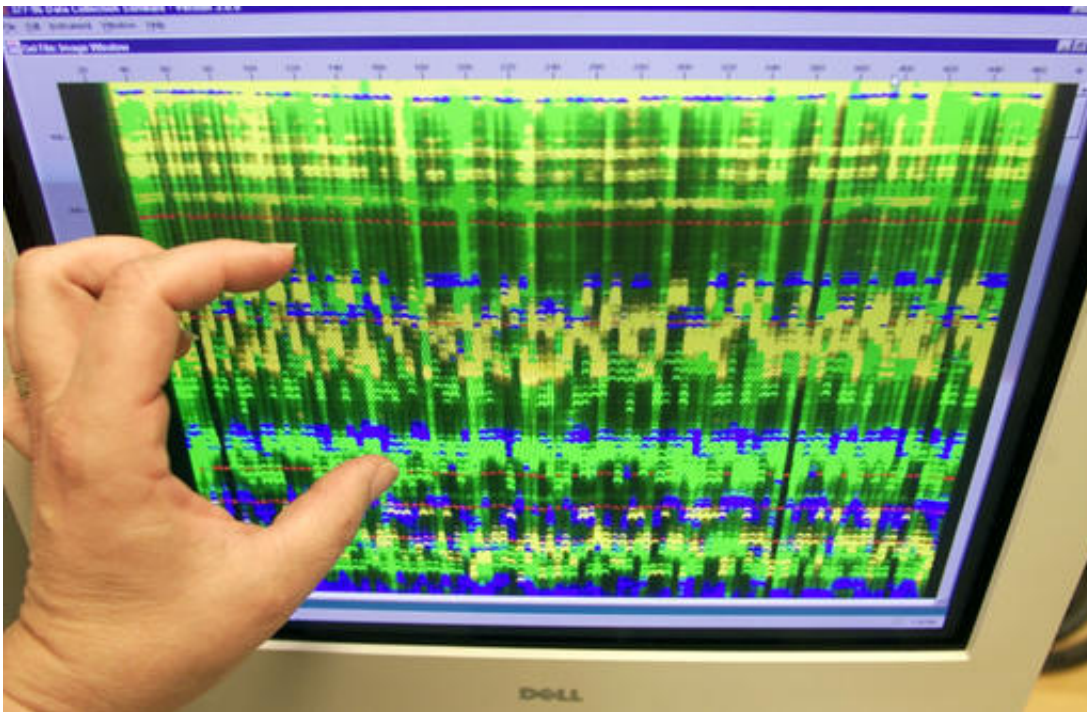


Scientists work toward storing digital information in DNA

July 23 2016, by Malcolm Ritter



This Wednesday, April 24, 2002 photo shows a gel image of the DNA of 96 horses displayed on a computer monitor at the UC Davis veterinary genetics lab in Davis, Calif. DNA is an information-storing molecule; the genes passed from generation to generation transmit the blueprints for creating the organism. (AP Photo/Eric Risberg)

Her computer, Karin Strauss says, contains her "digital attic"—a place where she stores that published math paper she wrote in high school, and computer science schoolwork from college.

She'd like to preserve the stuff "as long as I live, at least," says Strauss, 37. But computers must be replaced every few years, and each time she must copy the information over, "which is a little bit of a headache."

It would be much better, she says, if she could store it in DNA—the stuff our genes are made of.

Strauss, who works at Microsoft Research in Redmond, Washington, is working to make that sci-fi fantasy a reality.

She and other scientists are not focused in finding ways to stow high school projects or snapshots or other things an average person might accumulate, at least for now. Rather, they aim to help companies and institutions archive huge amounts of data for decades or centuries, at a time when the world is generating digital data faster than it can store it.

To understand her quest, it helps to know how companies, governments and other institutions store data now: For long-term storage it's typically disks or a specialized kind of tape, wound up in cartridges about three inches on a side and less than an inch thick. A single cartridge containing about half a mile of tape can hold the equivalent of about 46 million books of 200 pages apiece, and three times that much if the data lends itself to being compressed.

A tape cartridge can store data for about 30 years under ideal conditions, says Matt Starr, chief technology officer of Spectra Logic, which sells data-storage devices. But a more practical limit is 10 to 15 years, he says.

It's not that the data will disappear from the tape. A bigger problem is familiar to anybody who has come across an old eight-track tape or floppy disk and realized he no longer has a machine to play it. Technology moves on, and data can't be retrieved if the means to read it is no longer available, Starr says.

So for that and other reasons, long-term archiving requires repeatedly copying the data to new technologies.

Into this world comes the notion of DNA storage. DNA is by its essence an information-storing molecule; the genes we pass from generation to generation transmit the blueprints for creating the human body. That information is stored in strings of what's often called the four-letter DNA code. That really refers to sequences of four building blocks—abbreviated as A, C, T and G—found in the DNA molecule. Specific sequences give the body directions for creating particular proteins.

Digital devices, on the other hand, store information in a two-letter code that produces strings of ones and zeroes. A capital "A," for example, is 01000001.

Converting [digital information](#) to DNA involves translating between the two codes. In one lab, for example, a capital A can become ATATG. The idea is once that transformation is made, strings of DNA can be custom-made to carry the new code, and hence the information that code contains.

One selling point is durability. Scientists can recover and read DNA sequences from fossils of Neanderthals and even older life forms. So as a storage medium, "it could last thousands and thousands of years," says Luis Ceze of the University of Washington, who works with Microsoft on DNA data storage.

Advocates also stress that DNA crams information into very little space. Almost every cell of your body carries about six feet of it; that adds up to billions of miles in a single person. In terms of information storage, that compactness could mean storing all the publicly accessible data on the internet in a space the size of a shoebox, Ceze says.

In fact, all the digital information in the world might be stored in a load of whitish, powdery DNA that fits in space the size of a large van, says Nick Goldman of the European Bioinformatics Institute in Hinxton, England.

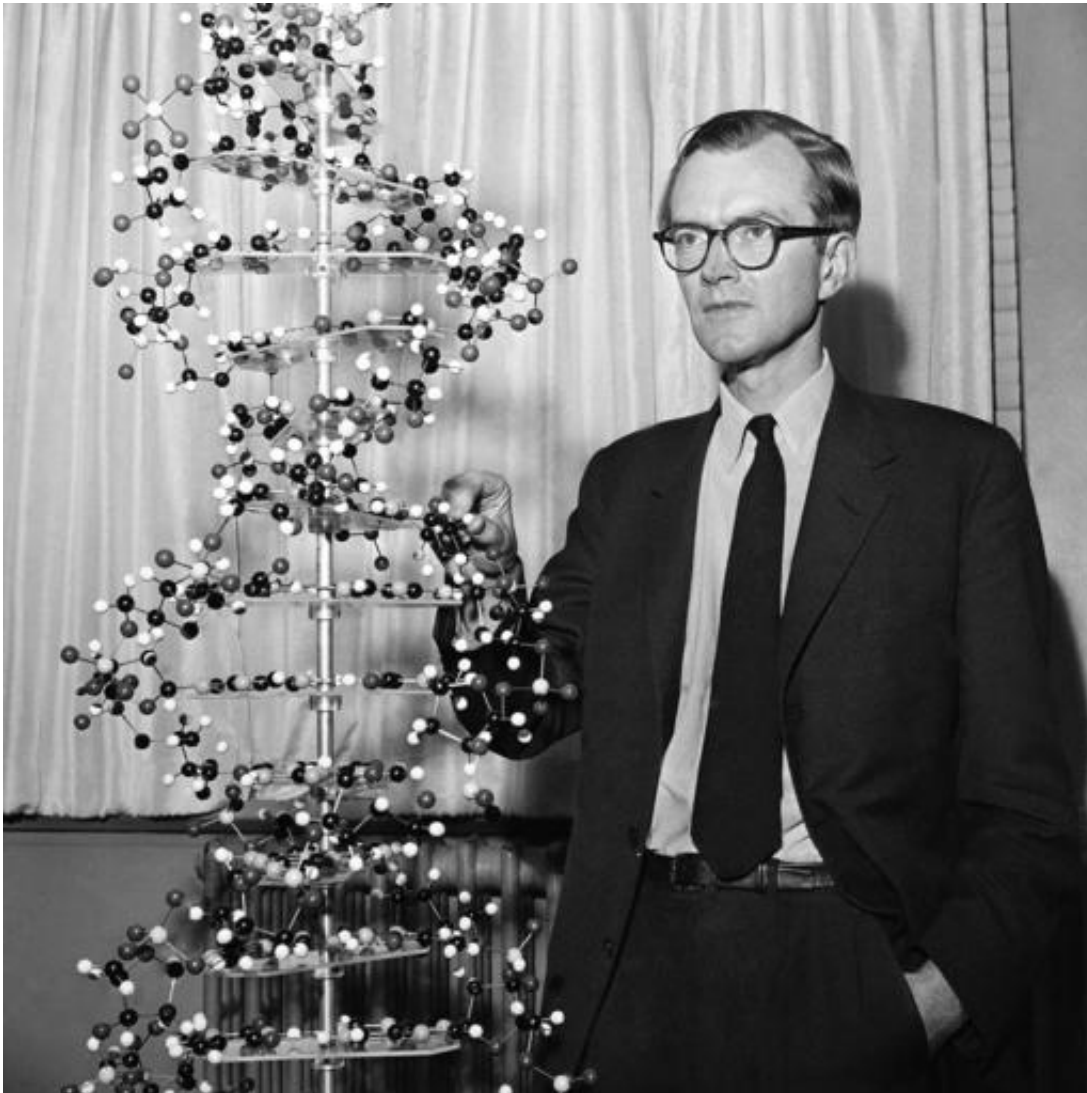
What's more, advocates say, DNA storage would avoid the problem of having to repeatedly copy stored information into new formats as the technology for reading it becomes outmoded.

"There's always going to be someone in the business of making a DNA reader because of the health care applications," Goldman says. "It's always something we're going to want to do quickly and inexpensively."

Getting the information into DNA takes some doing. Once scientists have converted the digital code into the 4-letter DNA code, they have to custom-make DNA. For some recent research Strauss and Ceze worked on, that involved creating about 10 million short strings of DNA.

Twist Bioscience of San Francisco used a machine to create the strings letter by letter, like snapping together Lego pieces to build a tower. The machine can build up to 1.6 million strings at a time.

Each string carried just a fragment of information from a digital file, plus a chemical tag to indicate what file the information came from.



In this Oct. 18, 1962 file photo, Dr. Maurice Hugh Frederick Wilkins, 46, of Greenwich, England, stands with a model of a DNA molecule during a news conference in the New York office of the Sloan-Kettering Institute for Cancer Research. Specific sequences of four building blocks—abbreviated as A, C, T and G—found in the DNA molecule give an organism directions for creating particular proteins. Wilkins shared the Nobel Prize for medicine with two other biochemists, Drs. Francis Harry Compton Crick and James Dewey Watson. (AP Photo/Anthony Camerano)

To read a file, scientists use the tags to assemble the relevant strings. A

standard lab machine can then reveal the sequence of DNA letters in each string.

Nobody is talking about replacing hard drives in consumer computers with DNA. For one thing, it takes too long to read the stored information. That's never going to be accomplished in seconds, says Ewan Birney, who works on DNA storage with Goldman at the bioinformatics institute.

But for valuable material like corporate records in long-term storage, "if it's worth it, you'll wait," says Goldman, who with Birney is talking to investors about setting up a company to offer DNA storage.

Sri Kosuri of the University of California Los Angeles, who has worked on DNA information storage but now largely moved on to other pursuits, says one challenge for making the technology practical is making it much cheaper.

Scientists custom-build fairly short strings DNA now for research, but scaling up enough to handle information storage in bulk would require a "mind-boggling" leap in output, Kosuri says. With current technology, that would be hugely expensive, he says.

George Church, a prominent Harvard genetics expert, agrees that cost is a big issue. But "I'm pretty optimistic it can be brought down" dramatically in a decade or less, says Church, who is in the process of starting a company to offer DNA storage methods.

For all the interest in the topic, it's worth noting that so far the amount of information that researchers have stored in DNA is relatively tiny.

Earlier this month, Microsoft announced that a team including Strauss and Ceze had stored a record 200 megabytes. The information included

100 books—one, fittingly, was "Great Expectations"—along with a brief video and many documents. But it was still less than 5 percent the capacity of an ordinary DVD.

Yet it's about nine times the mark reported just last month by Church, who says the announcement shows "how fast the field is moving."

Meanwhile, people involved with archiving digital data say their field views DNA as a possibility for the future, but not a cure-all.

"It's a very interesting and promising approach to the storage problem, but the storage problem is really only a very small part of digital preservation," says Cal Lee, a professor at the University of North Carolina's School of Information and Library Science.

It's true that society will probably always have devices to read DNA, so that gets around the problem of obsolete readers, he says. But that's not enough.

"If you just read the ones and zeroes, you don't know how to interpret it," Lee says.

For example, is that string a picture, text, a sound clip or a video? Do you still have the software to make sense of it?

What's more, the people in charge of keeping digital information want to check on it periodically to make sure it's still intact, and "I don't know how viable that is with DNA," says Euan Cochrane, digital preservation manager at the Yale University Library. It may mean fewer such check-ups, he says.

Cochrane, who describes his job as keeping information accessible "10 years to forever," says DNA looks interesting if its cost can be reduced

and scientists find ways to more quickly store and recover [information](#).

Starr says his data-storage device company hasn't taken a detailed look at DNA technology because it's too far in the future.

There are "always things out on the horizon that could store data for a very long time," he says. But the challenge of turning those ideas into a practical product "really trims the field down pretty quickly."

© 2016 The Associated Press. All rights reserved.

Citation: Scientists work toward storing digital information in DNA (2016, July 23) retrieved 20 March 2024 from <https://phys.org/news/2016-07-scientists-digital-dna.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--