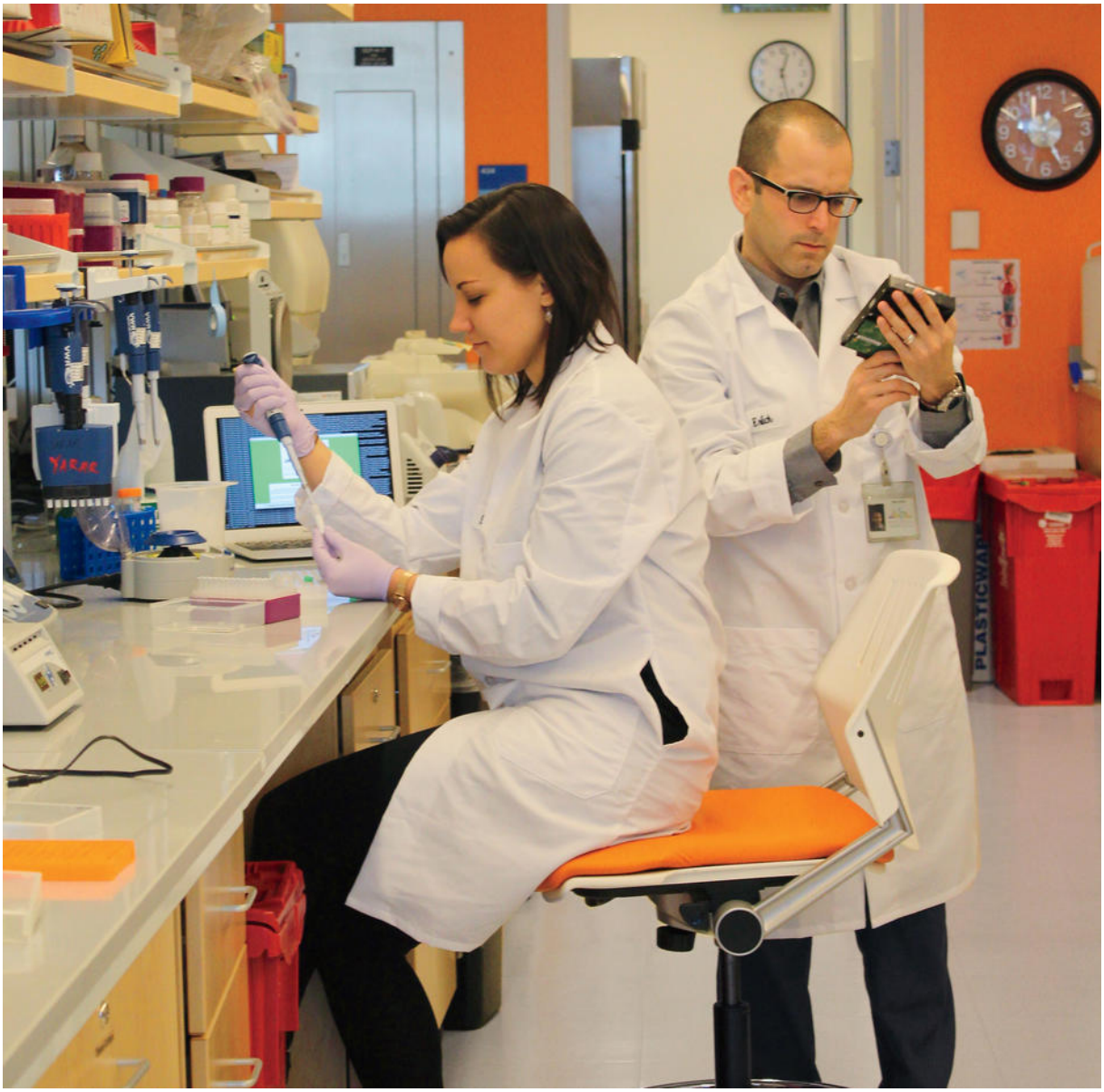
Science & Technology

The Future of Data Storage is in Our DNA

By
David J. Craig
|
Summer 2017



Dina Zielinski and Yaniv Erlich. Photo: New York Genome Center

As humanity creates more and more digital data, archiving the information on hard drives presents economic and ecological challenges. Today an ever-expanding network of large data-storage centers already accounts for more than 2 percent of all electricity consumption in the United States.

In an effort to develop cheaper and more sustainable storage methods, some scientists have begun experimenting with putting data onto nature's original hard drive: DNA. In the past five years, a number of research groups have shown that synthetic forms of DNA can be encoded with words, images, or music just as easily as with biological information.

Now Yaniv Erlich, an assistant professor of computer science at Columbia Engineering, and Dina Zielinski, a bioinformatics researcher at the New York Genome Center, have achieved a major breakthrough in this area, developing a technique that has enabled them to fit 60 percent more digital data onto a given strand of DNA than was previously possible. In a recent issue of the journal *Science*, the researchers describe how they managed to squeeze a trove of digital content — including a copy of the 1895 Lumière brothers film *Arrival of a Train at La Ciotat*, a full computer operating system, a $50 Amazon gift card, a computer virus, and a 1948 study by information theorist Claude Shannon — onto a speck of DNA so small that if it existed in a living organism, it would likely carry the blueprints for just a handful of proteins. They say that their technique could theoretically enable scientists to cram millions of megabytes of information onto a single *gram* of DNA.

"To the best of our knowledge, this is the highest-density storage device ever created," says Erlich.

Erlich and Zielinski's storage technique is also very reliable. The researchers say that even after they induced the DNA to make copies of itself, and then forced those copies to make copies, and so on, the resulting double helices were found to contain flawless replicas of the original data.

"We really tortured the content to see if there was anything we could do to make errors appear," Erlich says. "But each time we read the files back onto our computers, they worked perfectly."

Downloading data onto DNA is still too expensive for commercial use; it cost Erlich and Zielinski about $9,000 to store and retrieve theirs. But the researchers suspect

that if they and other scientists can continue to improve the efficiency with which they translate computer code, with its long strings of 0s and 1s, into the chemical language of DNA, made up of various combinations of the four nucleotides adenine (A), guanine (G), cytosine (C), and thymine (T), the strategy could eventually provide a cost-efficient option for archiving everything from Facebook posts to historical documents.

Read more from ***David J. Craig***



Guide to school abbreviations

All categories >
Read more from ***David J. Craig***