

Storing data in DNA brings nature into the digital universe

July 28 2017, by Luis Ceze And Karin Strauss



Credit: AI-generated image ([disclaimer](#))

Humanity is producing data at an unimaginable rate, to the point that storage technologies can't keep up. Every five years, the amount of data we're producing increases [10-fold](#), including photos and videos. Not all of it needs to be stored, but manufacturers of data storage aren't making hard drives and flash chips fast enough to [hold what we do want to keep](#).

Since we're not going to stop taking pictures and recording movies, we need to develop new ways to save them.

Over millennia, nature has evolved an incredible information [storage medium](#) – DNA. It evolved to store genetic information, blueprints for building proteins, but DNA can be used for many more purposes than just that. DNA is also much denser than modern storage media: The data on hundreds of thousands of DVDs could fit inside a [matchbox-size package of DNA](#). DNA is also much more durable – [lasting thousands of years](#) – than today's hard drives, which may last [years or decades](#). And while hard drive formats and connection standards become obsolete, DNA never will, at least so long as there's life.

The idea of storing digital data in DNA is [several decades old](#), but recent work from [Harvard](#) and the [European Bioinformatics Institute](#) showed that progress in modern DNA manipulation methods could make it both possible and practical today. Many research groups, including at the [ETH Zurich](#), the [University of Illinois at Urbana-Champaign](#) and [Columbia University](#) are working on this problem. Our [own group](#) at the University of Washington and Microsoft [holds the world record](#) for the amount of data successfully stored in and retrieved from DNA – 200 megabytes.

Preparing bits to become atoms

Traditional media like hard drives, thumb drives or DVDs store digital data by changing either the [magnetic](#), [electrical](#) or [optical properties](#) of a material to store 0s and 1s.

To store data in DNA, the concept is the same, but the process is different. DNA molecules are long sequences of smaller molecules, called nucleotides – adenine, cytosine, thymine and guanine, usually designated as A, C, T and G. Rather than creating sequences of 0s and

1s, as in electronic media, DNA storage uses sequences of the nucleotides.

There are several ways to do this, but the general idea is to assign [digital data](#) patterns to DNA nucleotides. For instance, 00 could be equivalent to A, 01 to C, 10 to T and 11 to G. To store a picture, for example, we start with its encoding as a digital file, like a JPEG. That file is, in essence, a long string of 0s and 1s. Let's say the first eight bits of the file are 01111000; we break them into pairs – 01 11 10 00 – which correspond to C-G-T-A. That's the order in which we join the nucleotides to form a DNA strand.

Digital computer files can be quite large – [even terabytes in size for large databases](#). But individual DNA strands have to be much shorter – holding only about 20 bytes each. That's because the longer a DNA strand is, the harder it is to build chemically.

So we need to break the data into smaller chunks, and add to each an indicator of where in the sequence it falls. When it's time to read the DNA-stored information, that indicator will ensure all the chunks of data stay in their proper order.

Now we have a plan for how to store the data. Next we have to actually do it.

Storing the data

After determining what order the letters should go in, the DNA sequences are manufactured letter by letter with chemical reactions. These reactions are driven by equipment that takes in bottles of A's, C's, G's and T's and mixes them in a liquid solution with other chemicals to control the reactions that specify the order of the physical DNA strands.

This process brings us another benefit of DNA storage: backup copies. Rather than making one strand at a time, the chemical reactions make many identical strands at once, before going on to make many copies of the next strand in the series.

Once the DNA strands are created, we need to protect them against damage from [humidity and light](#). So we dry them out and put them in a container that keeps them cold and blocks water and light.

But stored data are useful only if we can retrieve them later.

Reading the data back

To read the data back out of storage, we use a sequencing machine exactly like those used for analysis of [genomic DNA in cells](#). This identifies the molecules, generating a letter sequence per molecule, which we then decode into a binary sequence of 0s and 1s in order. This process can destroy the DNA as it is read – but that's where those backup copies come into play: There are many copies of each sequence.

And if the backup copies get depleted, it is easy to make duplicate copies to refill the storage – just as nature [copies DNA all the time](#).

At the moment, most DNA retrieval systems require reading all of the information stored in a particular container, even if we want only a small amount of it. This is like reading an entire hard drive's worth of information just to find one email message. We have developed techniques – based on [well-studied biochemistry methods](#) – that let us [identify and read](#) only the [specific pieces of information](#) a user needs to retrieve from DNA storage.

Remaining challenges

At present, DNA storage is experimental. Before it becomes commonplace, it needs to be completely automated, and the processes of both building DNA and reading it must be improved. They are both prone to error and relatively slow. For example, today's DNA synthesis lets us write a few [hundred bytes per second](#); a modern hard drive can write [hundreds of millions of bytes per second](#). An average iPhone photo would take several hours to store in DNA, though it takes less than a second to save on the phone or transfer to a computer.

These are significant challenges, but we are optimistic because all the relevant technologies are improving rapidly. Further, DNA data storage doesn't need the perfect accuracy that biology requires, so researchers are likely to find even cheaper and faster ways to store information in nature's oldest data storage system.

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: Storing data in DNA brings nature into the digital universe (2017, July 28) retrieved 26 April 2024 from <https://phys.org/news/2017-07-dna-nature-digital-universe.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> |
|--|