

Researchers break record for DNA data storage

July 8 2016, by Jennifer Langston



A depiction of the double helical structure of DNA. Its four coding units (A, T, C, G) are color-coded in pink, orange, purple and yellow. Credit: NHGRI

University of Washington and Microsoft researchers have broken what they believe is the world record for the amount of digital data successfully stored—and retrieved—in DNA molecules.

The team of computer scientists and electrical engineers encoded and decoded a video of the band OK Go (featuring the craziest Rube Goldberg machine ever), the Universal Declaration of Human Rights in more than 100 languages, the top 100 books of Project Gutenberg and the Crop Trust's seed database—among other things— all on strands of DNA.

Luis Ceze, the UW's Torode Family Career Development Professor of computer science and engineering and one of the project's lead researchers, expands on the latest news-making accomplishment from the UW Molecular Information Systems Lab:

Why are people interested in using DNA to store digital data?

LC: The world is producing data at an incredible rate, and storage technologies need to keep up. DNA is a remarkable storage molecule—it is millions of times denser than other storage media, it is incredibly durable (think millennia) and it never becomes obsolete. We humans, as DNA-based life forms, will always be interested in reading and writing DNA.

How quickly are we running out of room to warehouse all the data—from quirky cat videos to shopping preferences to essential medical records—the world is producing?

LC: Very quickly. Already today we can't store all data produced. Sure, a lot of that data might not be so useful, but the gap is only increasing. That is especially true of all the video and genomic data that will be produced over the next decade.

How much data did the UW-Microsoft research team store and retrieve in DNA strands and what have you learned?

LC: We stored 200MB of data. This experiment led to several important breakthroughs that improved our ability to manipulate more complex pools of synthetic DNA. It allowed us to better understand what kinds of errors crop up and how to deal with them.

Why choose OK Go's "This Too Shall Pass" video?

LC: We wanted to store something creative and in a modern format. HD video was a natural choice for format. And OK Go—being such a creative band—was a perfect fit. Also, there is an interesting connection between Rube Goldberg machines and molecular biology. Nature has produced incredible molecular machines, and when looked at closely enough might resemble a very complex but very reliable Rube Goldberg machine—without the soundtrack though!

How do you encode digital data—which is made up of 1s and 0s—in the building blocks of DNA?

LC: Interestingly, DNA already has a digital "flavor," as it has four bases and molecules that "stick" to each other in a very programmable way. So the first step in storing digital data into DNA is to map strings of 1s and 0s into strings of As, Cs, Gs and Ts. Next, the DNA sequences are

actually "manufactured" chemically, in a very parallel way. Our collaborator Twist Bioscience has a silicon-based DNA synthesis substrate that can make many different sequences in parallel. After the DNA molecules are manufactured, they are put in a test tube and dehydrated. And if protected from light and heat, they can last a long—and I mean very long—time.

How can you find and retrieve the files you're looking for?

LC: When one wants to read data, the DNA is re-suspended and read by a DNA sequencer, which determines what A, C, G, T letters comprise the molecules. From that, our algorithms recover the original [digital data](#). Despite being reliable, DNA writing and reading have errors, just like hard drives and electronic memories have errors, so we needed to develop error-correcting codes to reliably retrieve data. We also developed a method for "random access," which means you selectively read only the data you want and not the whole thing. We do that by borrowing from nature again and using DNA amplification—using [polymerase chain reactions](#) specifically—to only amplify the desired data.

What's next for the Molecular Information Systems Lab?

LC: There are still many challenges in making DNA storage mainstream. We will continue to focus on developing an end-to-end system and work with our Microsoft and Twist Bioscience collaborators to reduce the cost and increase the speed of writing and reading DNA.

Provided by University of Washington

Citation: Researchers break record for DNA data storage (2016, July 8) retrieved 19 April 2024 from <https://phys.org/news/2016-07-dna-storage.html>

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.