

Power of DNA to store information gets an upgrade

July 13 2020



Credit: CC0 Public Domain

A team of interdisciplinary researchers has discovered a new technique to store in DNA information—in this case "The Wizard of Oz," translated into Esperanto—with unprecedented accuracy and efficiency. The technique harnesses the information-storage capacity of intertwined strands of DNA to encode and retrieve information in a way that is both durable and compact.

The technique is described in a paper in this week's *Proceedings of the National Academy of Sciences*.

"The key breakthrough is an encoding algorithm that allows accurate retrieval of the information even when the DNA strands are partially damaged during storage," said Ilya Finkelstein, an associate professor of molecular biosciences and one of the authors of the study.

Humans are creating information at exponentially higher rates than we used to, contributing to the need for a way to store more information efficiently and in a way that will last a long time. Companies such as Google and Microsoft are among those exploring using DNA to store information.

"We need a way to store this data so that it is available when and where it's needed in a format that will be readable," said Stephen Jones, a research scientist who collaborated on the project with Finkelstein; Bill Press, a professor jointly appointed in computer science and integrative biology; and Ph.D. alumnus John Hawkins. "This idea takes advantage of what biology has been doing for billions of years: storing lots of information in a very small space that lasts a long time. DNA doesn't take up much space, it can be stored at room temperature, and it can last for hundreds of thousands of years."

DNA is about 5 million times more efficient than current storage methods. Put another way, a one milliliter droplet of DNA could store the same amount of information as two Walmarts full of data servers. And DNA doesn't require permanent cooling and hard disks that are prone to mechanical failures.

There's just one problem: DNA is prone to errors. And when a [genetic code](#) has errors, it's a lot different from when a computer code has errors. Errors in computer codes tend to show up as blank spots in the

code. Errors in DNA sequences show up as insertions or deletions. The problem there is that when something is deleted or added in DNA, the whole sequence shifts, with no blank spots to alert anyone.

Previously, when information was stored in DNA, the piece of information that needed to be saved, such as a paragraph from a novel, would be repeated 10 to 15 times. When the information was read, the repetitions would be compared to eliminate any insertions or deletions.

"We found a way to build the information more like a lattice," Jones said. "Each piece of information reinforces other pieces of information. That way, it only needs to be read once."

The language the researchers developed also avoids sections of DNA that are prone to errors or that are difficult to read. The parameters of the language can also change with the type of information that is being stored. For instance, a dropped word in a novel is not as big a deal as a dropped zero in a tax return.

To demonstrate information retrieval from degraded DNA, the team subjected its "Wizard of Oz" code to high temperatures and extreme humidity. Even though the DNA strands were damaged by these harsh conditions, all the information was still decoded successfully.

"We tried to tackle as many problems with the process as we could at the same time," said Hawkins, who recently was with UT's Oden Institute for Computational Engineering and Sciences. "What we ended up with is pretty remarkable."

More information: [DOI: 10.1073/pnas.2004821117](https://doi.org/10.1073/pnas.2004821117) William H. Press et al., "HEDGES error-correcting code for DNA storage corrects indels and allows sequence constraints," *PNAS* (2020).
www.pnas.org/cgi/doi/10.1073/pnas.2004821117

Provided by University of Texas at Austin

Citation: Power of DNA to store information gets an upgrade (2020, July 13) retrieved 26 April 2024 from <https://phys.org/news/2020-07-power-dna.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.