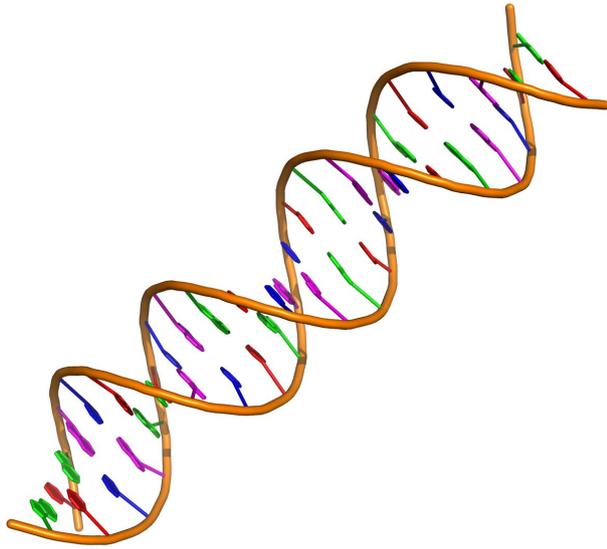


New twist on DNA data storage lets users preview stored files

10 June 2021



A double stranded DNA fragment. Credit: Vcpmartin/Wikimedia/ CC BY-SA 4.0

Researchers from North Carolina State University have turned a longstanding challenge in DNA data storage into a tool, using it to offer users previews of stored data files—such as thumbnail versions of image files.

DNA data storage is an attractive technology because it has the potential to store a tremendous amount of data in a small package, it can store that data for a long time, and it does so in an energy-efficient way. However, until now, it wasn't possible to preview the data in a file stored as DNA—if you wanted to know what a file was, you had to "open" the entire file.

"The advantage to our technique is that it is more efficient in terms of time and money," says Kyle Tomek, lead author of a paper on the work and a Ph.D. student at NC State. "If you are not sure

which file has the data you want, you don't have to sequence all of the DNA in all of the potential files. Instead, you can sequence much smaller portions of the DNA files to serve as previews."

Here's a quick overview of how this works.

Users "name" their [data files](#) by attaching sequences of DNA called primer-binding sequences to the ends of DNA strands that are storing information. To identify and extract a given file, most systems use polymerase chain reaction (PCR). Specifically, they use a small DNA primer that matches the corresponding primer-binding sequence to identify the DNA strands containing the file you want. The system then uses PCR to make lots of copies of the relevant DNA strands, then sequences the entire sample. Because the process makes numerous copies of the targeted DNA strands, the signal of the targeted strands is stronger than the rest of the sample, making it possible to identify the targeted DNA sequence and read the file.

However, one challenge that DNA data storage researchers have grappled with is that if two or more files have similar file names, the PCR will inadvertently copy pieces of multiple data files. As a result, users have to give files very distinct names to avoid getting messy data.

"At some point it occurred to us that we might be able to use these non-specific interactions as a tool, rather than viewing it as a problem," says Albert Keung, co-corresponding author of a paper on the work and an assistant professor of chemical and biomolecular engineering at NC State.

Specifically, the researchers developed a technique that makes use of similar file names to let them open either an entire file or a specific subset of that file. This works by using a specific naming convention when naming a file and a given subset of the file. They can choose whether to open the

entire file, or just the "preview" version, by manipulating several parameters of the PCR process: the temperature, the concentration of DNA in the sample, and the types and concentrations of reagents in the sample.

Provided by North Carolina State University

"Our technique makes the system more complex," says James Tuck, co-corresponding author of the paper and a professor of computer engineering at NC State. "This means that we have to be even more careful in managing both the file-naming conventions and the conditions of PCR. However, this makes the system both more data-efficient and substantially more user friendly."

The researchers demonstrated their technique by saving four large JPEG image files in DNA data storage and retrieving thumbnails of each file, as well as the full, high-resolution files in their entirety.

"Although we have only stored image files, this technology is broadly compatible with other file types. It also provides this new functionality without added cost," says Kevin Volkel, co-author of the work and a Ph.D. student at NC State.

The new "file preview" technique is also compatible with the DNA Enrichment and Nested Separation (DENSE) system that the researchers created to make DNA data storage more practical. DENSE effectively made DNA storage systems more scalable by introducing improved techniques for data file labeling and retrieval.

"We're currently looking for industry partners to help us explore the technology's commercial viability," Keung says.

The paper, "Promiscuous Molecules for Smarter File Operations in DNA-based Data Storage," will be published June 10 in *Nature Communications*. The paper was co-authored by Elaine Indermaur, a former undergraduate at NC State.

More information: "Promiscuous Molecules for Smarter File Operations in DNA-based Data Storage," *Nature Communications* (2021). [DOI: 10.1038/s41467-021-23669-w](https://doi.org/10.1038/s41467-021-23669-w)

APA citation: New twist on DNA data storage lets users preview stored files (2021, June 10) retrieved 3 December 2021 from <https://techxplore.com/news/2021-06-dna-storage-users-preview.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.