New approach to DNA data storage makes system more dynamic, scalable

June 12 2020

Credit: CC0 Public Domain

Researchers from North Carolina State University have developed a fundamentally new approach to DNA data storage systems, giving users the ability to read or modify data files without destroying them and making the systems easier to scale up for practical use.

"Most of the existing DNA data storage systems rely on polymerase

chain reaction (PCR) to access stored files, which is very efficient at copying information but presents some significant challenges," says Albert Keung, co-corresponding author of a paper on the work. "We've developed a system called Dynamic Operations and Reusable Information Storage, or DORIS, that doesn't rely on PCR. That has helped us address some of the key obstacles facing practical implementation of DNA data storage technologies." Keung is an assistant professor of chemical and biomolecular engineering at NC State.

DNA data storage systems have the potential to hold orders of magnitude more information than existing systems of comparable size. However, existing technologies have struggled to address a range of concerns related to practical implementation.

Current systems rely on sequences of DNA called primer-binding sequences that are added to the ends of DNA strands that store information. In short, the primer-binding sequence of DNA serves as a file name. When you want a given file, you retrieve the strands of DNA bearing that sequence.

Many of the practical barriers to DNA data storage technologies revolve around the use of PCR to retrieve stored data. Systems that rely on PCR have to drastically raise and lower the temperature of the stored genetic material in order to rip the double-stranded DNA apart and reveal the primer-binding sequence. This results in all of the DNA—the primer-binding sequences and the data-storage sequences—swimming free in a kind of genetic soup. Existing technologies can then sort through the soup to find, retrieve and copy the relevant DNA using PCR. The temperature swings are problematic for developing practical technologies, and the PCR technique itself gradually consumes—or uses up—the original version of the file that is being retrieved.

DORIS takes a different approach. Instead of using double-stranded DNA as a primer-binding sequence, DORIS uses an "overhang" that consists of a single-strand of DNA—like a tail that streams behind the double-stranded DNA that actually stores data. While traditional techniques required temperature fluctuations to rip open the DNA in order to find the relevant primer-binding sequences, using a single-stranded overhang means that DORIS can find the appropriate primer-binding sequences without disturbing the double-stranded DNA.

"In other words, DORIS can work at room temperature, making it much more feasible to develop DNA data management technologies that are viable in real-world scenarios," says James Tuck, co-corresponding author of the paper and a professor of electrical and computer engineering at NC State.

The other benefit of not having to rip apart the DNA strands is that the DNA sequence in the overhang can be the same as a sequence found in the double-stranded region of the data file itself. That's difficult to achieve in PCR-based systems without sacrificing information density—because the system wouldn't be able to differentiate between primer-binding sequences and data-storage sequences.

"DORIS allows us to significantly increase the information density of the system, and also makes it easier to scale up to handle really large databases," says Kevin Lin, first author of the paper and a Ph.D. student at NC State.

And once DORIS has identified the correct DNA sequence, it doesn't rely on PCR to make copies. Instead, DORIS transcribes the DNA to RNA, which is then reverse-transcribed back into DNA which the data-storage system can read. In other words, DORIS doesn't have to consume the original file in order to read it.

The single-stranded overhangs can also be modified, allowing users to rename files, delete files or "lock" them—effectively making them invisible to other users.

"We've developed a functional prototype of DORIS, so we know it works," Keung says. "We're now interested in scaling it up, speeding it up and putting it into a device that automates the process—making it user friendly."

The paper, "Dynamic and scalable DNA-based information storage," will be published June 12 in the journal *Nature Communications*.

More information: "Dynamic and scalable DNA-based information storage," *Nature Communications* (2020). DOI: 10.1038/s41467-020-16797-2

Provided by North Carolina State University