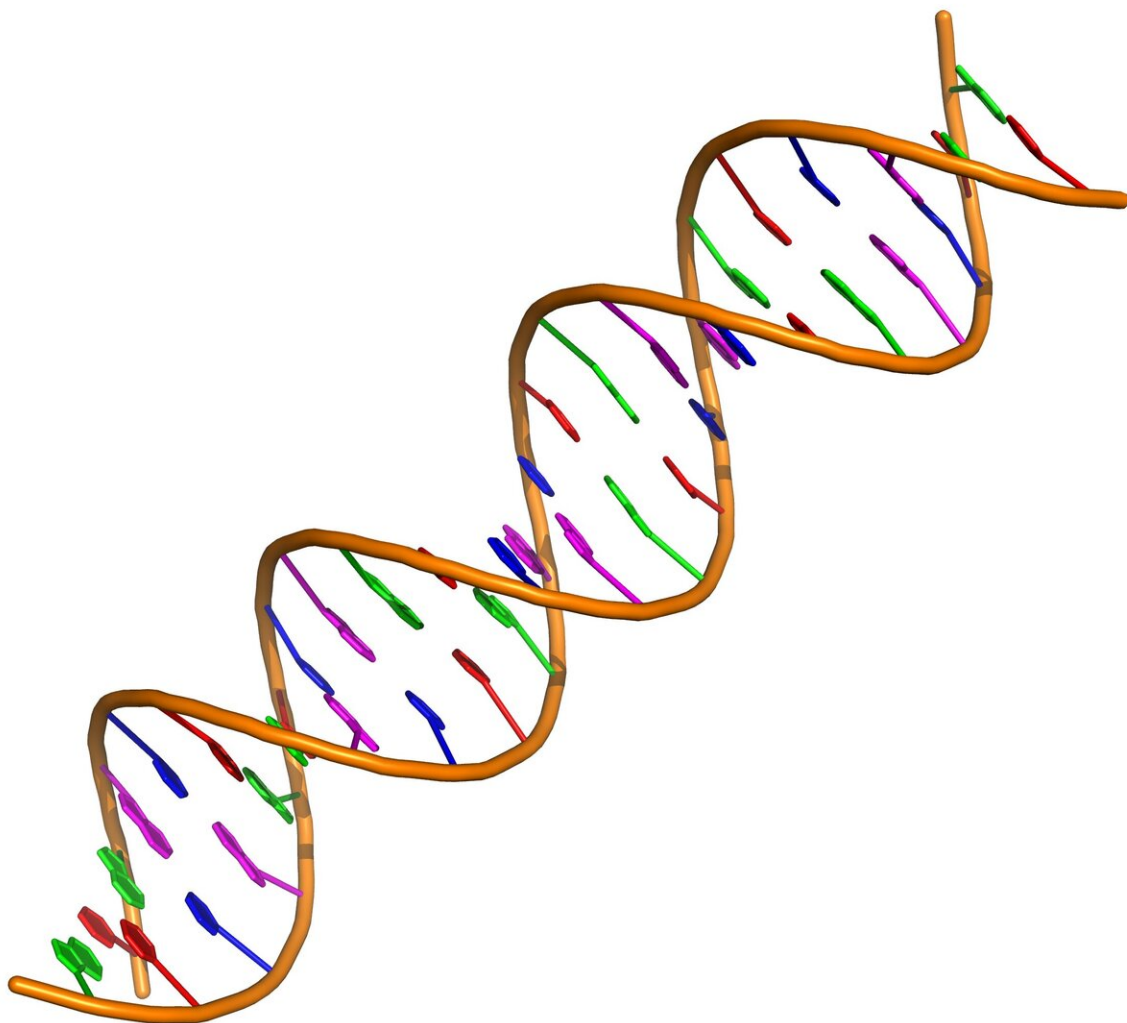# Dual-plasmid editing system improves DNA digital storage potential

August 5 2022

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

4.0

DNA-based information is a new interdisciplinary field linking information technology and biotechnology. The field hopes to meet the enormous need for long-term data storage by using DNA as an information storage medium. Despite DNA's promise of strong stability, high storage density and low maintenance cost, however, researchers face problems accurately rewriting digital information encoded in DNA sequences.

Generally, DNA data storage technology has two modes, i.e., the "in vitro hard disk mode" and the "in vivo CD mode." The primary advantage of the in vivo mode is its low-cost, reliable replication of chromosomal DNA by cell replication. Due to this characteristic, it can be used for rapid and low-cost data copy dissemination. Since encoded DNA sequences for some information contain a large number of repeats and the appearance of homopolymers, however, such information can only be "written" and "read," but cannot be accurately "rewritten."

To solve the rewriting problem, Prof. Liu Kai from the Department of Chemistry, Tsinghua University, Prof. LI Jingjing from the Changchun Institute of Applied Chemistry (CIAC) of the Chinese Academy of Sciences, and Prof. Chen Dong from Zhejiang University led a research team that recently developed a dual-plasmid editing system for accurately processing [digital information](link) in a microbial vector. Their findings were published in *Science Advances*.

The researchers established a dual-plasmid system in vivo using a rationally designed coding algorithm and an information editing tool. This dual-plasmid system is suitable for storing, reading and rewriting various types of information, including text, codebooks and images. It

fully explores the coding capability of DNA sequences without requiring any addressing indices or backup sequences. It is also compatible with various kinds of coding algorithms, thus enabling high coding efficiency. For example, the coding efficiency of the current system reaches 4.0 bits per nucleotide.

To achieve high efficiency as well as reliability in rewriting complex information stored in exogenous DNA sequences in vivo, a variety of CRISPR-associated proteins (Cas) and recombinase were used. The tools were guided by their corresponding CRISPR RNA (crRNA) to cleave a target locus in a DNA sequence so that the specific information could be addressed and rewritten. Because of the high specificity between complementary pairs of nucleic acid molecules, the information-encoded DNA sequences were accurately reconstructed by recombinase to encode new information. Due to optimizing the crRNA sequence, the information rewriting tool became highly adaptable to complex information, thus resulting in rewriting reliability of up to 94%, which is comparable to existing gene-editing systems.

The dual-plasmid system can serve as a universal platform for DNA-based information rewriting in vivo, thus offering a new strategy for information processing and target-specific rewriting of large and complicated data on a molecular level.

"We believe this strategy can also be applied in a living host with a larger genome, such as yeast, which would further pave the way for practical applications regarding big data storage," said Prof. Liu.

More information: Yangyi Liu et al, In vivo processing of digital information molecularly with targeted specificity and robust reliability, *Science Advances* (2022). DOI: 10.1126/sciadv.abo7415. www.science.org/doi/10.1126/sciadv.abo7415