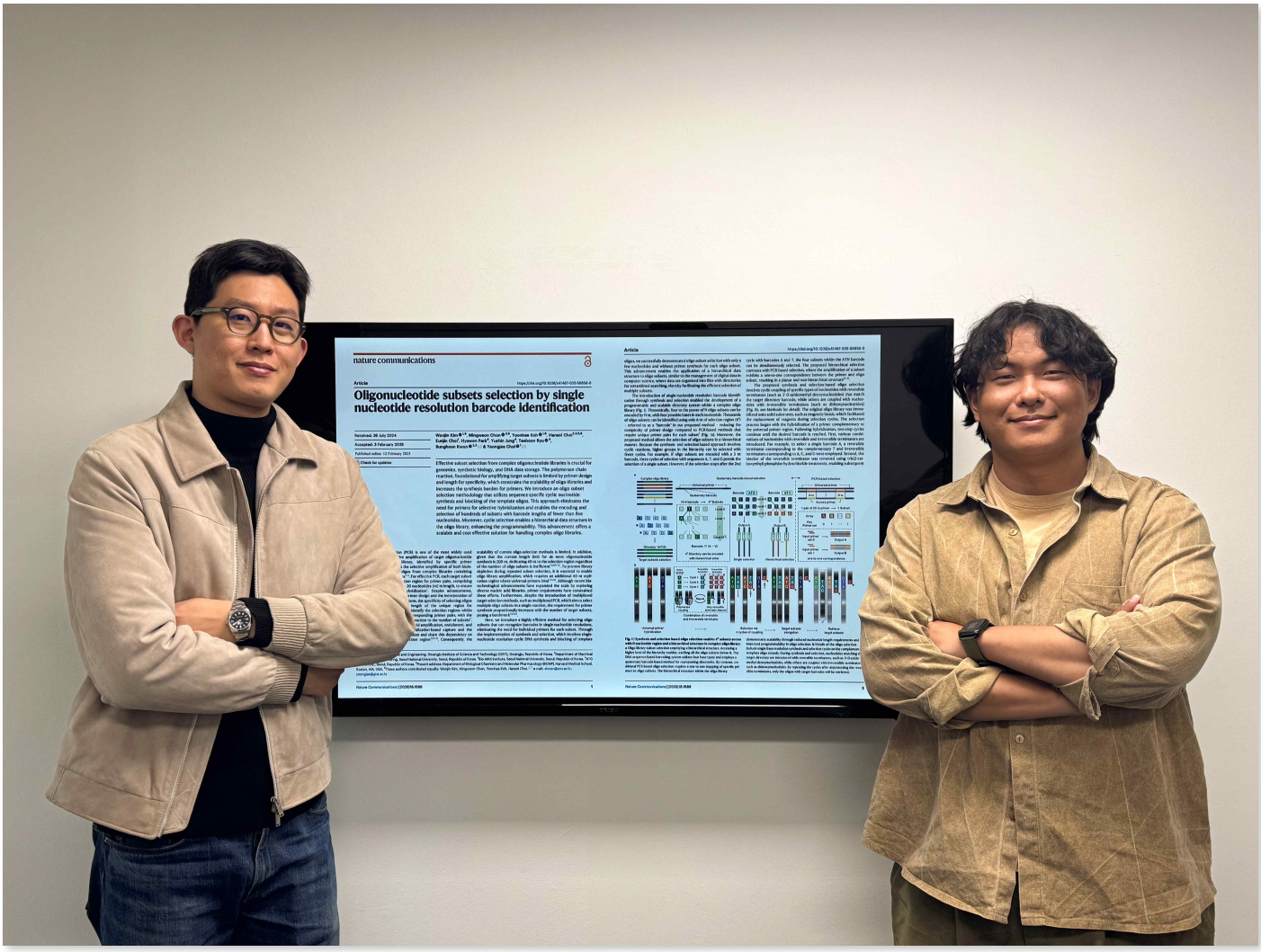
“Dream storage technology takes a big step toward reality”
GIST develops next-generation DNA storage technology, securing data center-level storage efficiency and scalability by freely accessing over 400 billion DNA files without primers

- GIST, Seoul National University, and ATG Lifetech Joint Research Team Develop ‘Cyclic DNA Synthesis and Selection’ Method… Access to Desired DNA File with Just 4 Bases, Data Integration 74 Million Times Higher than Existing Methods

- “An Important Breakthrough in Commercializing DNA Storage Systems Expected” published in the international journal 《Nature Communications》



▲ (From left) Professor Yeongjae Choi of the School of Materials Science and Engineering and Woojin Kim, a student in the combined master's and doctoral program

As DNA* application technology research is actively being conducted worldwide, semi-permanent and low-maintenance DNA-based storage methods are attracting attention as next-generation memory technologies.

A new DNA file access technology developed by a Korean research team is expected to dramatically increase data integration to the level of a ‘data center’ by overcoming the limitations of existing silicon semiconductor memory.

* DNA (deoxyribonucleic acid): A type of chemical substance that contains the genetic information of most living organisms (excluding some viruses).

The Gwangju Institute of Science and Technology (GIST, President Kichul Lim) announced that the research team of Professor Yeongjae Choi of the School of Materials Science and Engineering, Professor Sunghoon Kwon of Seoul National University, and the research team of AT&T Lifetech, Inc., have developed a 'cyclic DNA synthesis and selection' method*.

Using this technology, specific files within DNA data can be found more precisely and freely manipulated.

* cyclic DNA Synthesis and Selection: This is a method that repeatedly performs DNA synthesis and selection processes to select a specific DNA sequence. First, in the DNA synthesis step, only DNA containing the target sequence is continuously synthesized on a single-stranded template DNA, and the other sequences are stopped. Then, in the selection process, only the desired sequence is left and the rest is removed, and this process is repeated several times to finally leave only the specific DNA sequence. This allows for precise selection of the desired DNA sequence without a primer.

Existing DNA file access technologies such as PCR (polymerase chain reaction)* and hybridization capture require designing different primers* to amplify or physically separate specific DNA.

However, primers must contain at least 20 bases, and therefore have a structural limitation that requires additional allocation of a long sequence to recognize a specific DNA. In addition, as the types of DNA files to be distinguished increase, there was a problem that the cost of designing and synthesizing primers to distinguish them increased exponentially.

Therefore, a new biochemical method is needed that can efficiently identify and store DNA files without using primers, and can respond to various data sizes and complex structures.

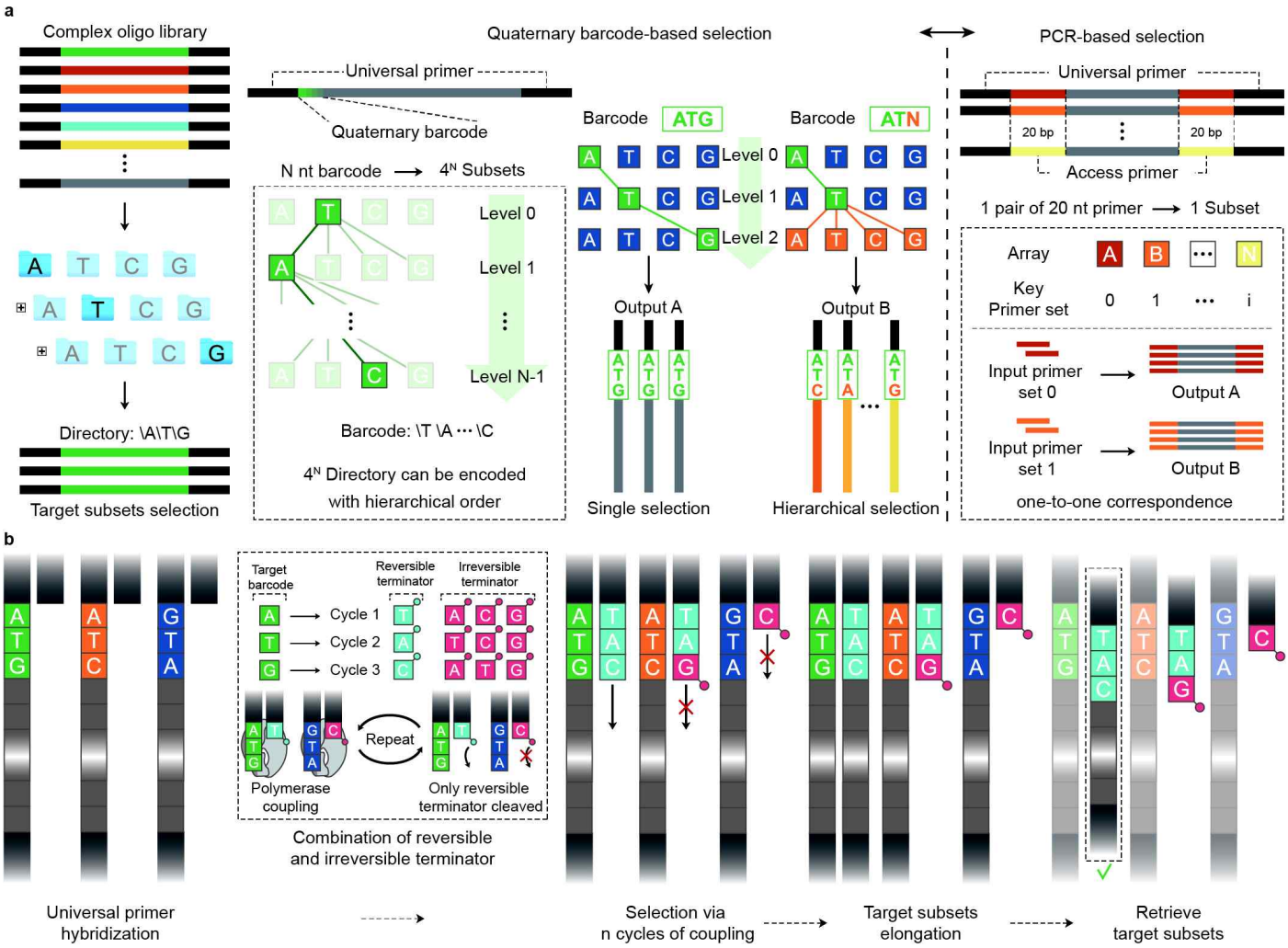
* PCR (polymerase chain reaction): A technology that amplifies a specific DNA sequence, allowing a large amount of desired DNA to be replicated in a short period of time. This process proceeds by repeating three steps: denaturation, annealing, and extension. First, heat is applied to separate (denature) the double helix of DNA and make it into a single strand. After that, a primer that can complementarily bind to a specific sequence binds (binds) to the DNA, and DNA polymerase synthesizes (extends) a new DNA strand using the primer as the starting point. By repeating these three steps, the amount of DNA increases exponentially with each cycle, and ultimately, the desired DNA sequence is amplified in large quantities.

* primer: A short DNA fragment, usually consisting of 20 bases, that serves to find a specific part during DNA replication or amplification. When a primer recognizes and binds to a specific sequence, the DNA replication process begins from that point.

The 'cyclic DNA synthesis and selection' technology developed by the research team was designed to search DNA files in a hierarchical structure (hierarchical selection method)* without primers by utilizing single-base-level barcodes.

This is a similar concept to the folder search method of a computer, and the cost is reduced by 10 times compared to the existing PCR method, and the access efficiency is improved by more than 3 times. In addition, the number of DNA files that can be distinguished has increased by at least 74 million times, and file replacement has become possible, such as deleting a specific DNA file and inserting a new DNA file.

* hierarchical selection method: It is a very efficient method in terms of data structure, like using a folder instead of handling files individually on a computer.



▲ Overview of cyclic DNA synthesis and selection technology. DNA libraries are structured hierarchically, so that when you access the upper level, you can easily reach all the DNA sets included in the lower level. The DNA base sequence-based barcoding system expresses directories in a quaternary (four-base) manner using four bases, and does not require a one-to-one correspondence with a specific primer as in the conventional PCR method. In addition, single-base level synthesis and selection cycles are performed on the complementary template DNA strand. Bases matching the target directory are added using a reversible terminator (e.g., 3'-O-azidomethyl deoxynucleotide), and non-target bases are combined using an irreversible terminator (e.g., dideoxynucleotide). After removing the irreversible terminator, the cycle is repeated to selectively recover only the DNA with the desired barcode.

In the conventional PCR method, a separate pair of primers (at least 20 bases) had to be designed and synthesized for each DNA file, but this technology can access a specific DNA file with only four bases.

At this time, it is highly scalable because it can theoretically distinguish 4n DNA files by adding the number of bases. For example, a barcode of 4 bases can distinguish 256 types of DNA files, and a barcode of 8 bases can distinguish 65,536 types of DNA files. In addition, using 20 bases, which is the length of a typical primer, it is theoretically possible to encode 415 billion subsets.

Professor Yeongjae Choi explained, “Through this study, we have proposed a new methodology that can access specific DNA files without primers, and by applying a hierarchical barcode system, we have overcome the limitations of the existing PCR method. If combined with barcode design optimization and automated systems in the future, it is expected to be an important breakthrough in the commercialization of DNA-based storage systems as a next-generation DNA file access technology.”

This research was supervised by Professor Yeongjae Choi of the School of Materials Science and Engineering at GIST, and conducted by Woojin Kim (integrated master's and doctoral program) and Yoonhae Koh (master's program), and was supported by the Ministry of Science and ICT's STEAM Research

Project (Future Network Convergence Technology Pioneer). The research results were published online in the international academic journal 《Nature Communications》 on February 12, 2025.