PCR'S NEXT WAVE

New applications are driving the need for faster, higher fidelity PCR. Nathan Blow looks at some clever modifications taking PCR to the next level.

Credit: Science Museum London/Science and Society Picture Library

DNA might just be the best data storage device on the planet—an optimal hard drive for both your body and, possibly, your computer. With four different bases for encoding information and billions of combinations of compact nucleotides all possible in a single solution, the potential exists to store large volumes of information within DNA. In recent years, researchers have described several ways to code information into DNA, but technical challenges associated with both writing and reading the data have limited enthusiasm.

In March 2017, another demonstration of data encoding in DNA was published in *Science* by Columbia University and New York Genome Center researchers Yaniv Erlich and Dina Zielinski (1). While any new approach to DNA storage is sure to elicit curiosity, Erlich and Zielinski's study was noteworthy for how much they were able to push DNA's storage limits.

Scientists use a metric known as Shannon capacity to measure how much information can be reliably encoded onto a device. Theoretically, the Shannon number for DNA is 2 bits per nucleotide, given that there are 4 nucleotides. In reality, however, this number has been much lower in experiments due to factors such as sequence limitations (homopolymer regions, for example, provide writing and reading problems) and technical efficiency issues (PCR amplification is not perfect and therefore reduces the potential information capacity). For these reasons previous studies have reported Shannon values of approximately 1.

Surprisingly, Erlich and Zielinski's latest system showed a Shannon value closer to 1.57 bits per nucleotide, nearly 60% higher than prior DNA coding schemes. But how did they do it? By changing the way they thought about traditional PCR.

Using PCR errors

Almost all DNA data encoding strategies have relied on PCR to either amplify or copy information stored within DNA. While

a robust way of making copies, this also means these approaches are subject to the shortcomings of PCR amplification, such as sensitivity issues, amplification bias, and polymerase errors. For many experiments—say amplifying a single region of interest—these shortcomings rarely create issues that cannot be overcome. However, in the case of large-scale or high-throughput experiments where fidelity is a key factor, such as when encoding and copying large amounts of data into DNA, PCR shortcomings rear their ugly heads more often.

Knowing this, Erlich and Zielinski sought a strategy for information encoding that would work with the PCR, rather than against it. What the researchers settled on was a system called fountain codes, which works for information coding in cases where the data is subject to dropout. A fountain coding approach, they argued, would enable more oligonucleotides of information to be generated and thereby increase redundancy, which should ultimately make for a DNA encoding process that is more robust in the face of PCR's shortcomings.

As it would turn out, fountain codes do indeed work very well. Using PCR and DNA sequencing on a fountain-encoded DNA library consisting of the code for an operating system, a movie, Shannon's information coding manuscript, and even an Amazon.com gift card, Erlich and Zielinski were able to decode the entire input file with zero errors. In fact, fountain coding was so robust that when they checked for errors by creating "deep copies" of the DNA data using 9 sequential PCR reactions—a potential 2.28 quadrillion copies of the file—they could still successfully decode the information stored in the DNA without error. According to Zielinski, creating "deep copies" is a very critical aspect of their work since it also demonstrates the potential to generate nearly unlimited volumes of data stored in DNA using PCR.

A fountain coding scheme was used by researchers to write multiple files—including movies and text—into DNA. Credit: D. Zielinski.

“What we did was to use the errors that can be made when performing PCR to test the limits of our DNA encoding and retrieval approach,” explains Zielinski. Since it is clear that PCR induces mutations, the fact that their encoding process using fountain codes could still decode DNA with no mistakes following repeated amplification cycles shows the potential of this approach.

Zielinski says that their method demonstrates that DNA data storage devices can be robust, in terms of both copying and propagating the DNA data libraries using PCR, and retrieving that information when needed.

Sensitive mistakes

PCR amplification of DNA is arguably one of the most important technology developments in the history of molecular biology, because without PCR, many commonly employed lab techniques would not be possible. It’s both robust and simple. Yet, there are still parts of the method that could be further refined—especially when it comes to sensitivity and error rate.

For researchers working in the cancer field, locating extremely

rare variants from complicated mixtures of DNA, proteins, RNA, and other biomolecules is often a daily task. But identifying ultra-rare variants of these molecules presents a huge challenge—it’s like finding a needle in a haystack, a process that can be confounded further by PCR errors. This is another area where developers are finding fertile ground for improving PCR methods and protocols.

Anders Stahlberg from the University of Gothenburg in Sweden, along with colleagues from Boston University School of Medicine and the Ontario Institute for Cancer Research, recently described a new multiplex PCR-based barcoding strategy that can be used for the detection of ultra-rare mutations by next-generation sequencing (2). According to Stahlberg, the method, which he and his colleagues call SIMSen-Seq (simple, multiplexed, PCR-based barcoding of DNA for sensitive mutation detection using sequencing), fills an existing gap for experimentalists.

“Our barcoding approach really acts as a bridge between the sensitivity of digital PCR and the throughput possible when working with next-generation sequencing,” says Stahlberg. As a quantitative approach, digital PCR is very sensitive for identifying rare genetic variants. The trouble is that digital PCR requires researchers know exactly what they are searching for—the technique is not a discovery tool or a way to seek out numerous rare variants at the same time. Applying next-generation sequencing to PCR-amplified DNA samples, on the other hand,

Researchers Yaniv Erlich (left) and Dina Zielinski (right) have created a new approach for encoding information into DNA that also exceeds the data storage limits of earlier methods. Credit: Y. Erlich and D. Zielinski.

Anders Stahlberg and colleagues have developed a barcoding approach to enhance the sensitivity of PCR for detecting rare genetic variants.
Credit: A. Stahlberg.

can be used to discover and analyze multiple rare variants, but this approach lacks the sensitivity of digital PCR and includes additional amplification steps when generating sequencing libraries that can lead to non-specific amplification products. It was with these shortcomings in mind that Stahlberg and his colleagues set out to build a better mousetrap—a way of achieving next-generation sequencing sensitivities closer to those of digital PCR.

Digital PCR works through separation—samples are separated into individual wells or reaction vessels where single molecules are independently amplified. In digital PCR, every molecule of DNA in a sample has the chance to be amplified; therefore, the number of molecules in a sample can be determined. To make something like this work using traditional PCR amplification in a single solution, however, researchers needed to develop a way of individually tagging, or barcoding, each amplified product such that all products could be sequenced and analyzed. While barcoding has been around for some time now, adding bar codes can further complicate PCR amplification. However, Stahlberg and his colleagues had an idea for their barcoding scheme to enhance sensitivity and eliminate errors at the same time—hide them.

“We actually hid our barcodes within a hairpin structure so that we would be able to control, using temperature, when a particular barcode opened or closed,” explains Stahlberg. The hairpin structure of Stahlberg’s barcodes means that they are closed (not able to be amplified) at the annealing temperature but open up

(able to be amplified) at the elongation temperature. In this way, the barcodes should not be able to act as primers during the PCR amplification step.

The barcoding and library production scheme worked. According to Stahlberg, the workflow takes around 3–4 hours to complete, and the data show that the sensitivity of this method is close to that of digital PCR—with the added benefit of being a discovery tool. He and his colleagues are now applying the approach to their tumor cell studies, and they hope the technique will find broader application among the many researchers looking for enhanced sensitivity in their own PCR experiments.

A fresh look at PCR

If the work of Erlich and Zielinski, and Stahlberg and his colleagues, shows us anything, it is that there are still ways to improve the laboratory workhorse known as PCR. But even as creative methods emerge to extend the range of PCR applications, others are noting that basic PCR reagents could also stand some enhancement.

In December 2016, a “Letter to the Editor” was published in *BioTechniques* from a research team in New Zealand led by Martin Kennedy. Kennedy’s letter documented a surprising finding from his lab—many hot-start polymerases that are being used to enhance PCR specificity were, in fact, working prior to their thermal activation. Hot-start polymerases are designed to only be active once a certain temperature is reached, thereby reducing the likelihood of non-specific amplification products being generated at lower temperatures.

Kennedy’s team saw a number of non-specific products when using different hot start polymerases, so they designed a specialized assay to test whether or not polymerase activity was present prior to thermal activation. Of 17 different hot-start polymerases that were tested, 12 were found to have polymerase activity prior to thermal activation. Kennedy and his colleagues’ take-home message is that researchers should carefully assess their polymerases prior to using and that the assay they detail can be used for this testing.

PCR is a foundational technology in the modern molecular biology laboratory—a robust technique that enables the exploration of DNA. But life science is entering a new era, where the experiments require greater sensitivity and higher output. In the face of these demands, even the most robust techniques are revealing their limitations. Fortunately, scientists often look at shortcomings as motivation to develop new methods. And for PCR, it’s never too late to teach an old dog some new tricks.

50 mL NOW AVAILABLE!
in addition to 10 and 25 mL

**LOSE THE WOBBLE
LOSE THE
DRIP**

a WIN for your work!

For more information go to www.vistalab.com/wobble-not

Wobble-not™
Low Insertion Force Serological Pipet

Never before seen
No Drip Design

The 2-tier/2-step design gives you 2X contact in your controller. You get more stability, meaning less wobble and consequently less dripping for less mess & less waste of time and reagents, and better accuracy, better performance, & better productivity.

VistaLab™
TECHNOLOGIES
Think Differently. Feel Better.

Don't mess with dripping. **GET FREE SAMPLES**
www.vistalab.com/wobble-not/#get-free-sample-pack

References

1. Erlich, Y. and D. Zielinski. 2017. DNA Fountain enables a robust and efficient storage architecture. *Science*. 355:950-954.
2. Stahlberg, A. et al. 2016. Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. *Nucleic Acids Res.* 44:e105.

Written by Nathan Blow, Ph.D.

BioTechniques 62:149-152 April 2017
doi: 10.2144/000114531