Researchers storing information securely in DNA

July 11 2016

Sandia National Laboratories bioengineers Marlene and George Bachand show off their new method for encrypting and storing information in DNA. Digital data storage degrades and can become obsolete and old-school books and paper require lots of space. Credit: Lonnie Anderson

Experiments at CERN's Large Hadron Collider generate 15 million gigabytes of data per year. That is a lot of digital data to inscribe on hard drives or beam up to the "cloud."

George Bachand, a Sandia National Laboratories bioengineer at the Center for Integrated Nanotechnologies, is exploring a better, more permanent method for encrypting and storing sensitive data: DNA. Compared to digital and analog information storage, DNA is more compact and durable and never becomes obsolete. Readable DNA was extracted from the 600,000-year-old remains of a horse found in the Yukon.

Tape- and disk-based data storage degrades and can become obsolete, requiring rewriting every decade or so. Cloud- or server-based storage requires a vast amount of electricity; in 2011 Google's server farms used enough electricity to power 200,000 U.S. homes. Furthermore, old-school methods require lots and lots of space. IBM estimated 1,000 gigabytes of information in book form would take up seven miles of bookshelves. In fact, Sandia recently completed a 15,000-square-foot building to store 35,000 boxes of inactive records and archival documents.

"Historically, the national laboratories and the U.S. government have a lot of highly secure information that they need to store long-term. I see this as a potentially robust way of storing classified information in the future to preserve it for multiple generations," said Bachand. "The key is how do you go from text to DNA and do that in a way that is safe and secure."

Bachand was inspired by the recording of all of Shakespeare's sonnets into 2.5 million base pairs of DNA—about half the genome of the tiny E. coli bacterium. Using this method, the group at the European Bioinformatics Institute could theoretically store 2.2 petabytes of

information—200 times the printed material in the Library of Congress—in one gram of DNA.

Marlene Bachand, a biological engineer at Sandia and George Bachand's spouse, added, "We are taking advantage of a biological component, DNA, and using its unique ability to encode huge amounts of data in an extremely small volume to develop DNA constructs that can be used to transmit and store vast amounts of encrypted data for security purposes."

The Bachands' project, funded by Sandia's Laboratory Directed Research & Development program, has successfully moved from the drawing board to letterhead. Using a practically unbreakable encryption key, the team has encoded an abridged version of a historical letter written by President Harry Truman into DNA. They then made the DNA, spotted it onto Sandia letterhead and mailed it—along with a conventional letter—around the country. After the letter's cross-country trip, the Bachands were able to extract the DNA out of the paper, amplify and sequence the DNA, and decode the message in about 24 hours at a cost of about $45.

Encrypting text into DNA and producing the message

To achieve this proof-of-principle, the first step was to develop the software to generate the encryption key and encrypt text into a DNA sequence. Andrew Gomez worked on this while he was an intern at Sandia; he is now at Senior Scientific, a nanomedicine company at the University of New Mexico's Science and Technology Park.

DNA is made up of four different bases, commonly referred to by their one letter abbreviations: A, C, G and T. Using a three-base code, exactly how living organisms store their information, 64 distinct characters—letters, spaces and punctuation—can be encoded, with room for redundancy.

The Bachands’ method of encrypting a message into DNA. Using a computer algorithm they can encrypt a message into a sequence of DNA. Then they chemically synthesize the DNA. The DNA can be read by DNA sequencing, and then translated and decoded using the same computer algorithm. Credit: Sandia National Laboratories

For example, spaces make up on average 15 to 20 percent of the characters in a text document, an encryption key could specify that TAG, TAA and TGA each code for "space" while GAA and CTC could code for "E." This would reduce the amount of repetition—technically challenging for making and reading DNA—and make brute-force hacking more difficult.

The team's first test was to encode a 180-character message, about the size of a tweet. Encoding the message into 550 bases was easy; actually making the DNA was hard.

"Our initial approach was very expensive, very time consuming and didn't work," said George Bachand. However, "there's a new technology that's come out and made the ability to take synthetic DNA, what are called gene blocks, and stitch them together into these artificial

chromosomes. These changes have just happened within the last few years, which has made it pretty extraordinary. Now it is possible to readily make these gene blocks right on the bench top and it can be done in large, production-scale pretty quickly."

Identifying potential national security applications

Since successfully encoding, making, reading and decoding the 180-character message and the 700-character Truman letter, the Bachands are now working on even longer test sequences. However, what the Bachands really want to do is move beyond tests and apply their technique to national security problems.

"We have achieved the proof-of-principle. Yes, it is possible. Now the big challenge for us is identifying the potential applications," said George Bachand. "Using DNA to store information is pretty cool, it's science-fiction-y, but the real question is it really good for anything? Can it really supplant any of the current technology and where we're headed in the future?"

Two possible applications the team has identified are storing historical classified documents and barcoding/watermarking electromechanical components, such as computer chips made in the Microsystems and Engineering Sciences Applications complex, Sandia's Department of Defense-certified fabrication facility, prior to storage.

George Bachand imagines encoding each component's history—when it was manufactured, the lot number, starting material, even the results of reliability tests—into DNA and spotting it onto the actual chip. Instead of having to find the serial number and look up that metadata in a digital or paper-based database, future engineers could swab the chip itself, sequence the DNA and get that information in a practically tamper-proof manner.

To test the feasibility, Marlene Bachand spotted lab equipment with a test message, and was able recover and decode the message, even after months of daily use and routine cleaning. DNA spotted onto electronic components and stored in cool, dark environments could be recoverable for hundreds of years.

Another, more straightforward application for the Bachands' DNA storage method would be for historical or rarely accessed classified documents. DNA requires much less maintenance than disk- or tape-based storage and doesn't need lots of electricity or tons of space like cloud- or paper-based storage. But conversion of paper documents into DNA requires the "cumbersome" process of scanning, encrypting, then synthesizing the DNA, admitted George Bachand. Making the DNA is the most expensive part of the process, but the cost has decreased substantially over the past few years and should continue to drop.

"I hope this project progresses and expands the biological scope and nature of projects here at Sandia. I believe the field of biomimicry has no boundaries. Given all of the issues with broken encryption and data breaches, this technology could potentially provide a path to address these timely and ever-increasing security problems," said Marlene Bachand.

Provided by Sandia National Laboratories

Citation: Researchers storing information securely in DNA (2016, July 11) retrieved 7 September 2024 from https://phys.org/news/2016-07-dna.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.