How Shakespeare and MLK Got Encoded in DNA

AP

David Wagner 4,067 Views Jan 24, 2013

Could the library of the future replace bookshelves with petri dishes? New research into the possibility of storing information in DNA has already preserved words by Shakespeare and Martin Luther King in genetic material. A project led by the European Bioinformatics Institute's Nick Goldman condensed the Bard's 154 sonnets into strands of DNA. They used a similar process to encode audio of Martin Luther King's "I Have a Dream" speech.

Here's how the process, outlined yesterday on the website of leading scientific journal *Nature*, works: The scientists took these writers' famous

words, encrypted them using a cipher that corresponds with DNA's four nucleic acids (A, C, G, or T), synthesized strands of DNA according to that code, and chilled the resulting samples in dark, dry conditions, where they should last for millennia. Goldman tells NPR's Adam Cole that one of our generation's biggest problems—organizing and storing the deluge of data we face every day—could be solved using DNA:

> The data we're being asked to be guardians of is growing exponentially. But our budgets are not growing exponentially ... We realized that DNA itself is a really efficient way of storing information.

This process shrinks information much more than existing formats like hard drives or magnetic tape. Or paper-bound books. Let's consider that a physical copy of *Shakespeare's Sonnets* from the Folger Shakespeare Library weighs 7 ounces. Project Gutenberg's digital version of the poems takes up 95 KB on your Kindle. That might seem pretty compact, but physical books and e-books are majorly inefficient storage methods when contrasted with genetic encoding. Shall we compare these to a strand of DNA? Goldman's team showed that they can fit the entire database of pioneering particle physics lab CERN (which holds approximately 90 petabytes of information) onto just 41 grams of DNA. In comparison, every sonnet Shakespeare ever wrote could fit on a mere speck of genetic material.

These findings aren't necessarily new—Harvard geneticist George Church was able to encode a book in DNA last summer. And some adventurous poets are even using DNA to encode new original works. In Canadian poet Christian Bök's four-line Xenotext, the couplet "Any style of life / is prim" is encoded in DNA that always spits out proteins reading "The faery is rosy / of glow." But even Church acknowledges the strides made by Goldman and his colleagues. "I think it’s a really important milestone," he told *Nature*'s Ed Yong. "We have a real field now."

Storing information in DNA is currently too expensive for standard commercial use. It costs about $12,400 to store a megabyte in DNA, and $220 to extract the information in readable form. But the expense is going down every year. "In 10 years, it's probably going to be about 100 times cheaper," Goldman told *The Wall Street Journal*'s Gautam Naik. "At that time, it probably becomes economically viable."

Want to add to this story? Let us know in comments or send an email to the author at dwagner@theatlantic.com. You can share ideas for stories on the Open Wire.

David Wagner

More at The Atlantic Wire