

From Hard Drives to Flash Drives to DNA Drives

M. Castillo, *Editor-in-Chief*

The word is now a virus.

William S. Burroughs, *The Ticket that Exploded*

Recently there has been another round of controversial news regarding genetically modified organisms (GMO). Perhaps the best known debate on these centers on corn. A recent French study showed severe kidney and liver abnormalities in rats that were fed this corn for up to 2 years.¹ Immediately afterward, Russia banned the use of this seed and the corn it produces. Because other studies have not confirmed this finding, the American media immediately released news stories stating that the French study was flawed and unscientific and that it represented just another round of propaganda by individuals who oppose GMO and the companies that produce the seeds (which are mostly American).² Salmon, with growth hormones that have been altered so that they not only grow faster but never stop growing, has also been in the news. Salmon is the third most-eaten seafood in the United States according to the National Fisheries Institute, and most of it is flown in from Chile, so growing enough of it here to feed Americans may actually be a good thing for the environment, even if its genes have been modified.³ All of these situations involve inserting or altering a specific gene in plant or animal deoxyribonucleic acid (DNA); thus, the genetic material in those organisms still serves its original purpose. However, what happens if we take our DNA, reconfigure it, and use it for something completely different from that for which it is intended? Cutting-edge genetic engineers are now synthesizing DNA so that it contains information much like a computer hard drive or solid memory chips. The capacity of DNA as a storage medium is staggering: All of the information contained on the entire Internet would fit into a device smaller than 1 cubic inch!

As our need for high-capacity information storage continues to increase, several researchers have begun to explore the possibility of using DNA for this purpose.⁴ The very fabric of life uses a binary code, but instead of the 1s and 0s computers use, the code in our DNA is composed of 4 letters: A, G, C, and T (adenine, guanine, cytosine, and thymine), which are paired into 2 nucleotide bases: A-T and G-C (hence a form of binary code). By changing the order of these 2 base nucleotide pairs, one can encode all different types of information in the same way a computer does by changing the order of 1s and 0s. Each nucleotide may encode 2 bits of information, and 1 g of single-stranded DNA can store 455 exabytes. One exabyte is equal to 1000 petabytes; 1 petabyte is equal to 1 quadrillion bytes, and so on. What this means is that in 1 g of single-stranded DNA, one can potentially store the equivalent

to 250 million DVDs! Computer chips are “planar” storage devices (obvious from their shape). One way to improve the capacity of a computer chip is to put several layers of circuits in it (making it 2D), but because DNA is 3D, it offers much more space. Memory cards are said to be reliable for up to 5 years after their initial use, but DNA-encoded information remains stable and readable for millennia.⁵ For purposes of timeless storage, DNA may be dried and then protected from water and oxygen, which gives it a nearly infinite stability.

DNA information storage is not new. It has been around since 1988, and one of the first successful projects came from the J. Craig Venter Institute, a nonprofit genomics research organization with facilities in 3 different US states. These investigators were able to encode 7920 bits into DNA.⁶ (Pridefully, in a synthetic cell, they encoded their names, 3 literary citations, and the address of an Internet site [Table].) Newer DNA-synthesizing techniques can alter the way base nucleotide pairs are formed, making it easier to encode information and thereafter read it. As mentioned previously, traditionally base pairs are A-T and G-C (remember that nucleotides are measured in pairs because DNA is usually double-stranded). Thus, the number of base pairs is equal to the number of nucleotides in 1 DNA strand. The problem with using the natural sequence of nucleotide base pairs for information encoding is that the G-C pair can be difficult to subsequently read. Therefore, new techniques use novel base pairs: A-C and G-T, which are easy to manufacture and thereafter interpret. With these 2 new base pairs, one also has a binary code: A-C for 0 and G-T for 1. At present, assembling long stable strands of DNA is difficult, so information needs to be parceled in smaller data blocks of DNA called “oligonucleotides” (by comparison, the human genome contains about 3 billion base pairs, so it is a very long strand and the amount of information that it contains is astonishing).

In a recent experiment, Church et al⁷ took one of their own books (nearly 54,000-words-long, including 11 images) and used a computer to convert it into a bit stream (they initially thought about encoding *Moby Dick*). They encoded all of the bits of the book into 159 oligonucleotides, each also containing information as to its general position within the text. The encoded DNA was then amplified by polymerase chain reaction* (PCR), and in this way, its base pairs could be assessed, read, and interpreted (similar techniques were used to map the human genome). During the entire process of writing, amplifying, and reading 5.27 megabits of information, only 10 bit errors occurred, a testament to how incredibly exact this technology is. Church et al were able to store in DNA 600 times more information than was previously possible. As amazing as this seems, one must add to it the fact that this technique used only in vitro procedures, avoiding the controversies of cloning and live genetic manipulations, and it was 100,000 times cheaper than other previous versions.

Synthetic DNA is exempt from the National Institutes of Health usage guidelines and is available to all with the means to

<http://dx.doi.org/10.3174/ajnr.A3482>

Works previously stored in DNA^a

| Year of Experiment | Usage Description | Specific Contents |
|--------------------|-------------------|---|
| 1988 | Art | Microvenus image ^b |
| 1998 | Text | Text from the Bible, Genesis |
| 2001 | Text | Parts from a book by Dickens |
| 2003 | Text | Parts from "It's a Small World," the main song of a musical boat ride from Walt Disney ^c |
| 2005 | Text | "Tomten" a poem by Viktor Rydberg ^d |
| 2010 | Watermark | Watermark of a synthetic genome ^e |

^a Adapted from Church et al.⁴

^b Image of the external female genitalia representing female Earth.

^c Encrypted into *Deinococcus Radiodurans*, a bacterium extremely resistant to inhospitable environments. Information resistant to the effects of a nuclear holocaust could be saved in similar cells.

^d A true example of "living poetry." For other similar projects, I suggest that you Google "Project Xenotext."

^e Watermarking a cell designed to contain information may help us keep track of it.

manufacture it. The cost of DNA synthesis drops 12-fold per year compared with that of newer electronic media (1.6-fold per year); thus, it is becoming widely available. For example, synthesizing a strand of DNA containing 100 million base pairs cost US \$10,000 in 2001 but only 10 cents today. Synthesizing and reading DNA for information-storage purposes will require 6–8 orders-of-magnitude improvement. Although this amount of improvement is significant, it will soon become a reality as handheld DNA sequencers become widely available and inexpensive. As the need to store untold amounts of information becomes more pressing, newer DNA-related technologies will be discovered and become less expensive.

In the supporting data from their article, Church et al.⁷ also bring up some safety and ethical concerns with regard to their experiment. They state that the DNA fragments they used to encode their book are "unlikely" to replicate themselves or encode anything else that could be biologically active. They do not discard the possibility that if this DNA were left out in the wild, it could get incorporated into a living organism. This last observation seems unlikely because cells tend to expel DNA that is not theirs. However, what would happen if an organism incorporates this foreign DNA has not even been a matter of speculation. Could a cell produce proteins hitherto unknown? Will that cell die? It certainly will not help us improve our individual knowledge because our bodies lack mechanisms with which to read this DNA and move its information to our brains. Ninety-eight percent of our DNA is now considered to be "genetic junk" (that is, DNA with no apparent function), so perhaps the day will come when we can use this space to encode into each human cell our history and accumulated knowledge.

All of our knowledge placed into highly resistant and self-replicating cells sent out to space in miniships may be the best way to explore the possibility of other civilizations existing far away from ours. Security and defense agencies have also considered DNA storage as a means of encryption. This technique was inspired by the World War II microdot technique of Germany, in which an entire page of information was photographed and reduced to the size of the dot at end of this sentence. DNA microdots can be hidden in general genetic material with their locations known only to those who know the primers marking the beginning and end of their specific DNA segments, which can then be resolved and read with PCR.⁸ Therefore, information could cross

borders in cells and not be subject to Internet counterespionage, and if the person carrying the information is detained, the site harboring the information would be impossible to detect.

Some say that all science fiction eventually becomes reality, and certainly DNA information storage must have sounded like science fiction just a few years ago. In Frank Herbert's novel *Dune* (Clinton Book Company, 1965), spaceships are able to navigate only because their control systems know at all times the positions of all celestial bodies. This tremendous amount of information is not saved in a computer but rather in mutated humans (the Guild Navigators), each controlling a spaceship. The Navigators can do this because their DNA contains all of the information needed for space travel. If a human being has more than 10 trillion cells, it does not seem far-fetched that his or her DNA could contain all of the information in the universe.

Update

Since I wrote this Perspectives, investigators at the European Bioinformatics Institute have found a new and different way to encode information into DNA. Dr. King's "I Have a Dream" speech, a photo, a PDF of Watson and Crick's seminal article, and all of Shakespeare's sonnets were encoded using it. The new method allows for multiple copies of this special DNA to be accurately manufactured. The authors expect their product to last over 10,000 years if kept dry, cold, and dark. Because storing information in DNA is easier than reading it, they suggest that DNA may be the ideal method for keeping information that does not need to be frequently accessed and thus ideal for libraries and government records. Please see the article in *Nature* by Goldman et al.⁹

*In this situation, PCR is easy to use because the makeup of the DNA strand that needs to be amplified is known. Primers that start the reaction can be easily chosen, and specific zones may be amplified and then read.

REFERENCES

1. Seralini GE, Clair E, Mesnage R, et al. **Long term toxicity of Roundup herbicide and a Roundup-tolerant genetically modified maize.** *Food Chem Toxicol* 2012;50:4221–31
2. Zhu Y, He X, Luo Y, et al. **A 90-day feeding study of glyphosate-tolerant maize with the G2-aroA gene in Sprague-Dawley rats.** *Food Chem Toxicol* 2012;51:280–87
3. **No Drop at the Top of NFI's Top 10 Most Popular List.** <http://www.aboutseafood.com/press/press-releases/no-drop-top-nfi-s-top-10-most-popular-list>. Accessed December 7, 2012
4. Church GM, Gao Y, Kosuri S. **Next-generation digital information storage in DNA.** *Science* 2012;337:1628
5. Bonnet J, Colotte M, Coudy D, et al. **Chain and conformation stability of solid-state DNA: implications of room temperature storage.** *Nucleic Acids Res* 2010;38:531–46
6. Gibson DG, Glass JI, Lartigue C, et al. **Creation of bacterial cell controlled by a chemically synthesized genome.** *Science* 2010;329:52–56
7. Church GM, Gao Y, Kosuri S. **Next-generation digital information storage in DNA** (supplementary material). <http://www.sciencemag.org/content/suppl/2012/08/15/science.1226355.DC1/Church.SM.pdf>. Accessed December 7, 2012
8. Clelland CT, Risca V, Bancroft C. **Hidden messages in DNA microdots.** *Nature* 1999;399:533–34
9. Goldman N, Bertone P, Chen S, et al. **Towards practical, high-capacity, low-maintenance information storage in synthesized DNA.** <http://www.nature.com/nature/journal/vaop/ncurrent/full/nature11875.html>. Accessed February 5, 2012