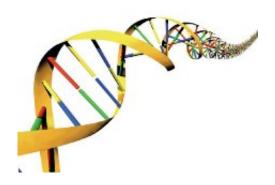# DNA used to encode a book and other digital information

August 17 2012, by Lin Edwards

(Phys.org) -- A team of researchers in the US has successfully encoded a 5.27 megabit book using DNA microchips, and they then read the book using DNA sequencing. Their experiments show that DNA could be used for long-term storage of digital information.

George Church and Sriram Kosuri of Harvard's Wyss Institute for Biologically Inspired Engineering, and colleagues, encoded Church's book "Regenesis" of around 53,400 words into DNA sequences, along with 11 images in JPG format and a JavaScript program. This is 1,000 times more data than has been encoded in DNA previously.

DNA is made up of nucleotides, and in theory at least each nucleotide can be used to encode two bits of data. This means that the density is a

massive 1 million gigabits per cubic millimeter, and only four grams of DNA could theoretically store all the digital data created annually. This is much denser than digital storage media such as flash drives, and more stable, since the DNA sequences could be read thousands of years after they were encoded.

The experiment's success lay in the strategy of encoding the data in short sequences of DNA rather than long ones, and this reduced the difficulty and cost of writing and reading the data. Dr Kosuri said the process was analogous to storing data on a hard drive, where data is written in small blocks called sectors.

They first converted the book, program and images to HTML and then translated this into a sequence of 5.27 million 0s and 1s, and these 5.27 megabits were then sequenced into sections of nucleotides 96 bits long using one DNA nucleotide for one bit. The nucleotide bases A and C encoded for 0, while G and T encoded for 1. Each block also contained a 19 bit address to encode the block's place in the overall sequence. Multiple copies of each block were synthesized to help in error correction.

After the book and other information was encoded into the DNA, drops of DNA were attached to microarray chips for storage. The chips were kept at 4°C for three months and then dissolved and sequenced. Each copy of each block of nucleotides was sequenced up to 3,000 times so that a consensus could be reached. In this way they reduced the bit errors in the 5.27 megabits to just 10.

The procedure, described in a paper in the journal *Science*, cannot be used for rewritable data but could be used for very long-term storage of data. One advantage of using DNA is that a much greater density of information can be stored, but another major advantage is that DNA is a biological molecule that will always be able to be read biologically

without special equipment such as CD or DVD players that can quickly become obsolete.

The main disadvantage of this system is that at the moment the technologies used to synthesize and sequence DNA are far too expensive for it to be a practical system for everyday use. Another problem is that while DNA has been sequenced from sources such as mummies thousands of years old, the DNA tends to be fragmented, and work needs to be done on improving the stability of DNA over centuries and longer.

ABSTRACT
Digital information is accumulating at an astounding rate, straining our ability to store and archive it. DNA is among the most dense and stable information media known. The development of new technologies in both DNA synthesis and sequencing make DNA an increasingly feasible digital storage medium. Here, we develop a strategy to encode arbitrary digital information in DNA, write a 5.27-megabit book using DNA microchips, and read the book using next-generation DNA sequencing.

© 2012 Phys.org