**Name: Abu Butt**

**Class: Csc 59866, Fall 2017**

**Professor: Abbe Mowshowitz**

**Topic: DNA Digital Data Storage**

**Introduction:**

Data storage and retrieval is inevitable and its preservation problem is looming over our information network. The demand for storing more and more data is increasing day by day. So to find new solutions to the issues of digital data storage, new technologies and principles are in a state of innovative experimentation throughout the world. DNA digital data storage refers to any process to store digital data in the base sequence of DNA. There are several reasons to use DNA as the storage medium. Its storage density and small size, something which would take conventional media roughly 2 million times that volume for the same amount of information.

**History:**

The idea and the general consideration about the possibility of recording, storage and retrieval of information on DNA molecules were originally made by *Mikhail Neiman* and were published in 1964-65 in the Radiotekhnika journal. Then there were several researches which have been critically reviewed in storing digital data on DNA. The idea of digital storage in DNA was first indirectly implemented in 1999 by Clell and Risca, Bancroft. They succeeded in storing encoded words in short DNA strands. Among early examples of DNA data storage, in 2007 a device was created at the University of Arizona, using addressing molecules to encode mismatch sites within a DNA strand. On August 16, 2012, the *journal Science* published research by George Church and colleagues at Harvard University, in which DNA was encoded with digital information that included HTML draft of a 53,400 words book written by the lead researches, eleven JPEG images and one JavaScript program. Furthermore, in March 2017, Yaniv Erlich and Dina Zielinski of Columbia University and the New York Genome Center published a method known as DNA Fountain, which allows perfect retrieval (~85%) of information from a density of petabytes per gram of DNA.

**DNA Digital Data Storage:**

DNA data storage is the process of encoding and decoding binary data onto and from synthesize strands of DNA (deoxyribonucleic acid). This technology uses artificial DNA made using commercially available oligonucleotide synthesis machine for storage and DNA sequencing machines for retrieval. To store a binary digital file as DNA, the individual bits are converted from 1 and 0 to the letters A, C, G, T. These are the 4 main compounds of DNA: adenine, cytosine, guanine, thymine. To recover the data, the same sequences, A, C, G, T representing the DNA molecules are decoded back into the original sequences of bits 1 and 0.

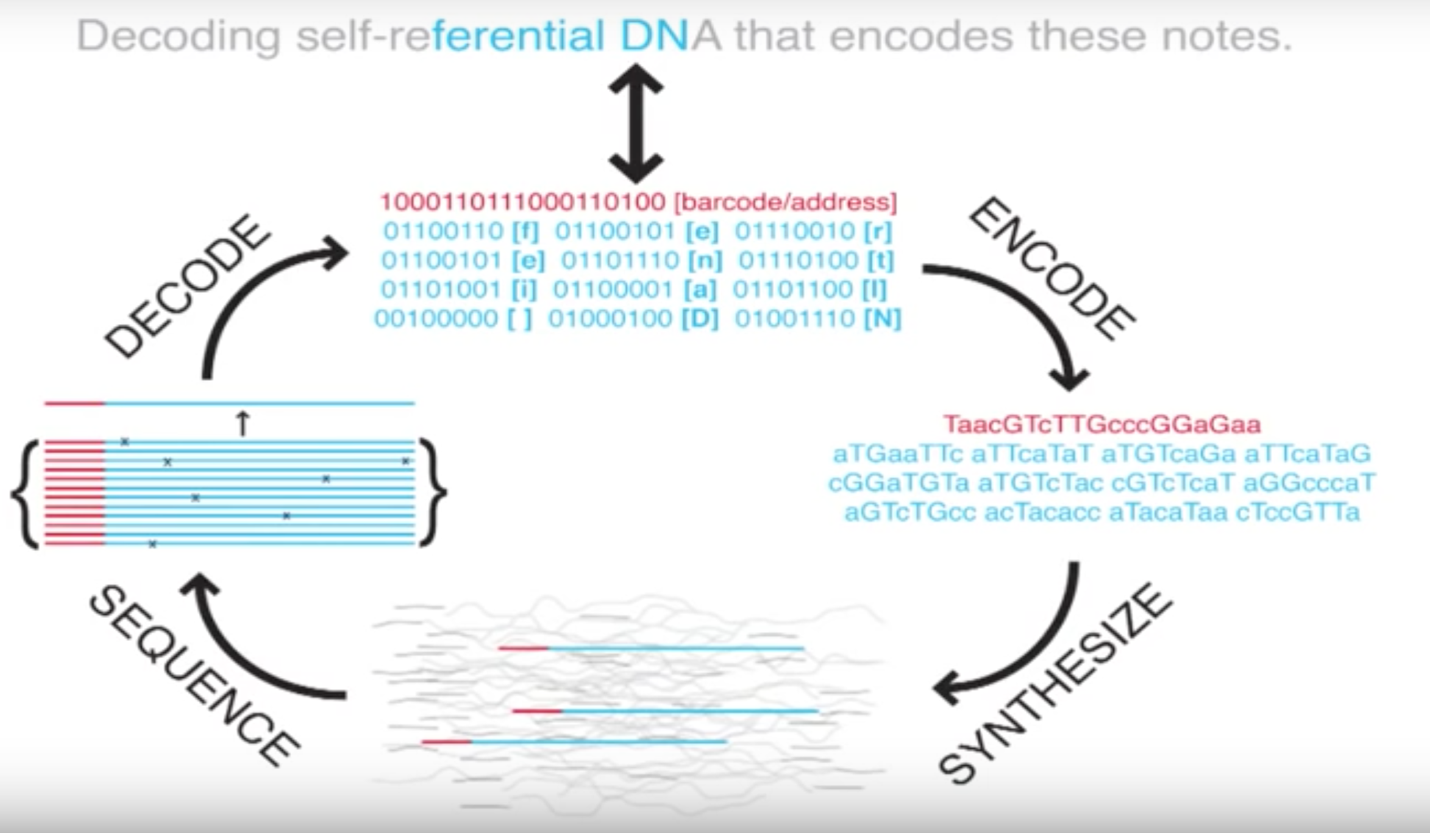


Figure 1. Overall Process

**DNA Digital Data Storage Advantages:**

Advantage of DNA digital storage is its longevity. Because DNA Molecules can survive for thousands of years, a digital archive encoded in this form could be recovered by people for many generations to come. The longevity might resolve the troubling prospect of our digital age being lost to history because of relative impermanence of optical, magnetic and electronic media. DNA can maintain its integrity without any power supply. Also, its small size and weight make it easy to store and transport. In addition, DNA is less susceptible to technical failure compare to the other electronic devices. Currently it was reported that in 1 gram of DNA 215 petabytes could be stored. At this size the entire world’s data could fit in a teaspoon. Data in DNA is stored in a volumetric fashion (using Adenine, Thymine, Guanine, Cytosine bases), which gives access to more storage options unlike present mediums which stores data in a linear model.

**DNA Digital Data Storage vs Computer Data Storage:**

Storing information in DNA differs from computer RAM in some pretty significant ways. Most notable is speed. The main reason of what makes RAM, RAM is the at it is easy-access system, which is also quick access system. Which allows it to hold data the computer might need at an instant’s notice, and make it available on those timescales. On the other hand, DNA is significantly harder and slower to read than conventional computer transistors, meaning in terms of access speed it’s less RAM-like than your average SSD or spinning magnetic hard drive. The other disadvantage of DNA digital data storage could be that DNA storage has a process of slow encoding speed and high cost. Since DNA has a slow process, as it needs to be sequenced in order to retrieve the large amount data.

**Remaining Challenges (Ethical or Unethical Issues):**

Considering all these major finding, it is inevitable that DNA would become a universal archival medium one day. But it presents several challenges, some due to its own physical composition, while some due to our technological ineptness to unleash its full potential at present. The overall process of encoding, amplifying, sequencing, restricting and decoding takes significant more time than their conventional counterparts. Consequently, DNA is unlikely to compete with optical, magnetic or quantum formats in the foreseeable future. At the risk of ignoring the benefits, the problems of the genetic testing have been widely discussed. This issue rises an ethical issue. Which includes whether this technology will benefit human life or not. Many clinical laboratories are archival stores for human tissues and very few have formal policies and procedures for the use of this material. One of the potential ethical issue is that this technology poses high risk in cryptography since DNA can encrypt computer DNA viruses. In addition, it can be used to encode and transport select vital agents that could be harmful to human or other organisms. The major disadvantage that could lead to an ethical issue is that DNA for data storage is the possible unanticipated effects on cell development and health. Another major challenge for practical DNA based information storage is the difficulty of synthesizing long sequences of DNA to a specified design.

**Conclusion:**

Data storage in DNA is no more confined to science fiction but us being realized and improvised at very promising rates by research teams all over the world. This idea has received positive criticisms from the public as can be inferred from their responses on the different science website. Similar all revolutions in technology, DNA-based data storage technology must face major challenges to realize its full potential. It is however, inevitable that DNA would be invariably used for archival purpose for its sheer density, robustness, stability, and energy efficiency. Digital Data Storage in DNA technology shows immense progress, since reading and writing it, is advancing ten times every year unlike Electronic Technology which is improving roughly 1.5 times a year. Furthermore, it will fuel research to look for alternative materials for information storage and to aid in realizing the need for a universal medium for data. Overall, this technology is here to stay ad could transform the way we have ever looked at DNA and computing as totally different entities.

**References:**

* Greene, J. (2017). Is DNA the Future of Data Storage. [online] WSJ. Available at: https://www.wsj.com/articles/is-dna-the-future-of-data-storage-1477405351 [Accessed 4 Oct. 2017].
* Meiling, Brittany. San Diego Business Journal. (2017). Data Storage Solution May Be in the DNA. [online] Available at: https://www.sdbj.com/news/2017/aug/24/data-storage-solution-may-be-dna/ [Accessed 4 Oct. 2017].
* Templeton, G. (2017). *How DNA data storage works - ExtremeTech*. [online] ExtremeTech. Available at: https://www.extremetech.com/extreme/231343-how-dna-data-storage-works-as-scientists-create-the-first-dna-ram [Accessed 11 Oct. 2017].
* <http://www.wsj.com/articles/is-dna-the-future-of-data-storage-1477405351>
* <https://en.wikipedia.org/wiki/DNA_digital_data_storage>
* <https://www.scientificamerican.com/article/tech-turns-to-biology-as-data-storage-needs-explode/>