RESEARCH ARTICLE

Deoxyribonucleic Acid as a Tool for Digital Information Storage: An Overview

Ashish C Patel[1], CG Joshi[2]

Abstract

Current data storage technologies cannot keep pace longer with exponentially growing amounts of data through the extensive use of social networking photos and media, etc. The "digital world" with 4.4 zettabytes in 2013 has predicted it to reach 44 zettabytes by 2020. From the past 30 years, scientists and researchers have been trying to develop a robust way of storing data on a medium which is dense and ever-lasting and found DNA as the most promising storage medium. Unlike existing storage devices, DNA requires no maintenance, except the need to store at a cool and dark place. DNA has a small size with high density; just 1 gram of dry DNA can store about 455 exabytes of data. DNA stores the informations using four bases, viz., A, T, G, and C, while CDs, hard disks and other devices stores the information using 0's and 1's on the spiral tracks. In the DNA based storage, after binarization of digital file into the binary codes, encoding and decoding are important steps in DNA based storage system. Once the digital file is encoded, the next step is to synthesize arbitrary single-strand DNA sequences and that can be stored in the deep freeze until use. When there is a need for information to be recovered, it can be done using DNA sequencing. New generation sequencing (NGS) capable of producing sequences with very high throughput at a much lower cost about less than 0.1 USD for one MB of data than the first sequencing technologies. Post-sequencing processing includes alignment of all reads using multiple sequence alignment (MSA) algorithms to obtain different consensus sequences. The consensus sequence is decoded as the reversal of the encoding process. Most prior DNA data storage efforts sequenced and decoded the entire amount of stored digital information with no random access, but nowadays it has become possible to extract selective files (e.g., retrieving only required image from a collection) from a DNA pool using PCR-based random access. Various scientists successfully stored up to 110 zettabytes data in one gram of DNA. In the future, with an efficient encoding, error corrections, cheaper DNA synthesis, and sequencing, DNA based storage will become a practical solution for storage of exponentially growing digital data.

Keywords: Byte, Decoding, Digital information, Encoding, PCR, Sequencing.

Ind J of Vet Sci and Biotech (2019): 10.21887/ijvsbt.15.1.1

Introduction

Social networking and the internet are producing huge data in the form of digital information. In 2013, the total amount of digital information worldwide was 4.4 zettabytes and is predicted to reach 44 zettabytes by 2020 (Bornholt *et al.,* 2016). Certain biological information, *viz.,* DNA, RNA, protein, etc. are also responsible for the generation of huge data. International Nucleotide Sequence Database Collaboration (INSDC) reported the assembled/ annotated portion of the sequence data maintained by the INSDC grew from 1.432 trillion bases in August 2015 to 2.650 trillion bases in August 2017. The growth rate over these two years was 85% just short of a doubling. During the same period, the read archive grew by 233% with the addition of 3000 trillion bases. The space required to store a single copy of the reads increased from 1.5 to 3.2 Petabytes, an increase of 113%, storage efficiency increased due to storing a greater fraction of submitted data in aligned and compressed format. As the data increases, the current data storage technology would not be enough to store data in the future as data is growing every day. In 2011, International data corporation (IDC) estimated that the global information storage capacity is 264 EB (Yim *et al.,* 2014). The demand for storing more and more data is increasing day by day, which urges the data scientists to strive for the development of better data storage medium.

The journey of data storage began from punched cards to magnetic tapes, drums, films, gramophone records, floppies, etc. Data storage has in the present scenario extended to optical discs including CDs, DVDs, blu-ray discs to portable hard drives and USB flash drives. Recently in 2017 (https://www.sony.net/SonyInfo), storage device as large as available of 330 TB capacity and is the densest form of storage available

[1]Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, AAU, Anand, Gujrat, India

[2]Gujarat Biotechnology Research Center (GBRC), GOG, Gandhinagar, Gujrat, India

Corresponding Author: Ashish C Patel, Department of Animal Genetics and Breeding, College of Veterinary Science and Animal Husbandry, AAU, Anand, Gujrat, India, e-mail: aashishvet@gmail.com

How to cite this article: Patel, A. C., & Joshi, C. G. (2019). Deoxyribonucleic Acid as a Tool for Digital Information Storage: An Overview. Ind J Vet Sci and Biotech, 15(1):1-8.

Source of support: Nil

Conflict of interest: None

Submitted: 30/3/2019 **Accepted:** 29/4/2019 **Published:** 27/7/2019

Table 1: Milestones of DNA as a tool for storage of text file and storage of digital information

1964–1965: Mikhail Neiman first time gave concept & possibilities of DNA based technology for data storage
1988: Craig Venter successfully encoded 7920 bits into DNA
1998: Dicken successfully encoded text from the Bible
2005: Viktor Rydberg successfully encoded text "Tomten" a poem successfully
2012: DNA as a tool for digital information storage
2012: George Church encoded and stored digital information of 53,400 words of the book, eleven JPG images and one JavaScript program in DNA and recovered information back successfully.
2013: Goldman *et al*. encoded and stored PDF, JPEG and MP3 files of 757 kB in DNA and recovered information back successfully
2015: Grass *et al*. encoded and stored 83 kB of a text file using error correction codes and recovered information back successfully
2017: Erlich and Zielinski developed a strategy to store and retrieve DNA information with a maximum of information per nucleotide and recovered information back successfully
2018: Organik *et al*. developed random access in large-scale DNA data storage

about 201 GB/in^2 (https://www.sony.net/SonyInfo). But all of these techniques are prone to obsolescence and decay. Moreover, they are made up of silicon, and the other non-biodegradable materials pollute the environment and would be exhaust one day. In 2005, the SD card was available with the maximum storage capacity of 128 MB of data while in 2014, it was available with 128 GB indicated, the storage facilities were increased 1000 times more within ten years (www.sandisk.com/home/memory-card). In spite of the improvement in data storage, to store zettabytes of data, millions of units are being required. So, it needs to increase storage density along with durability. Data loss can occur due to damage of hard drive, accidental deletion of files without backup or by poor handling of the optical disk (Laddha and Honwadkar, 2016). 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. Scientists and Researchers from different parts of the world have been testing to develop a robust way of storing data on a medium that is dense, universal, non-obsolete, ever-lasting and enduring and they found DNA as a potential medium for digital data storage. DNA is an attractive possibility because it is extremely dense and long-lasting, over 500 years (Church *et al.,* 2012).

History of digital information storage devices

In the past, the man had tried to find a way to store information, but nowadays people are using CD-ROM, USB, DVD, etc. After the 17th century, storage devices were invented in the form of the punch card, punch tape, phonograph, magnetic tape, magnetic drum, etc. In 1956, IBM invented the hard disk with the storage capacity of 5 MB. The first memory disk called the floppy disk was invented by Alan Shugart at IBM in 1971. It was considered as a revolutionary device for transporting data from one computer to another. Majority of information storage technologies were introduced between 1980 and 2010. At the beginning of the 1980s, the first optical devices, the CD and the CD ROM were released in the market. In 2003, the first blue-laser-based disc, the blue-rays disc was released. Several other "versions" in the form of the storage capacity of the DVDs and HD-DVD have been released and modified to store more and more data and to gain faster access. In 2006, the concept of cloud storage was introduced, where shared resources, data, and information are provided to computers and other devices on-demand. In 2012, an era of DNA based storage technology has emerged to store digital information. The milestones of DNA as a tool for storage of digital information are presented in Table 1.

Why dna as a digital information storage medium?

The current technologies for storage of the digital information are successfully used, but all existing storage technologies have certain drawbacks as they are prone to damage from magnetic fields, high temperature, moisture and all are releasing high amounts of heat energy. The most important drawback of current storage technologies is prone to obsolesce as technology advanced (Limbachiya and Gupta, 2015). So, synthetic DNA sequences have long been considered a potential medium for digital data storage. DNA is an attractive possibility because it is extremely dense, with a theoretical limit above 1 EB/mm^3 (eight orders of magnitude denser than tape), high storage capacity and long-lasting, with an observed half-life of over 500 years in harsh environments. DNA-based storage also has the benefit of everlasting relevance, as it will remain readable in the future too. Followings are the comparisons of different storage media in the form of their storage capacity and life expectancy (Table 2).

Table 2: Comparison of different storage media (Beck and Yampolskiy, 2015)

Type	*Life expectancy*	*Capacity*
DNA	Millions of year	108 TB/ 1 gram
Hard disk	~ 10 years	Up to 12 TB
SSD	> 10 years	Up to 4 TB
CD	~ 10 years	700 MB
DVD	< 10 years	Up to 17 GB
Blu–ray disc	30 to 100 years	25 GB (single layer)
USB Flash drive	~ 10 years	Up to 1 TB

How digital information can be stored in DNA?

First information is converted into digital form, if not. The data is converted into binary codes, this process is called as binarization, and these binary codes are further converted into A, T, G, C sequence with the help of a device, and this process of conversion is called as DNA encoding. Then further, these ATGC sequences are synthesized into DNA molecules and to read this information back from DNA, sequencing of DNA fragments is done by using various algorithms. After sequencing various decoding techniques are used to decode the data from it. The diagrammatical presentation of steps of DNA based storage system is presented in Figure 1.

Encoding and decoding of the digital file is an important step in DNA based storage system. The general encoding is carried out after binarization of digital file to obtain the binary codes. There are various encoding approaches proposed for DNA based information storage systems. *viz*., Clelland's Coding Scheme, Wong's Coding Scheme, DNA-Crypt coding scheme, ASCII Coding Scheme, Yachie's Coding Scheme, Arita's Coding Scheme, Coding Scheme based on the Huffman Code, Comma Code and Alternating Code, Church encoding model and Goldman encoding model. Among all the Coding Scheme, Huffman coding scheme and non-linear ternary coding scheme are considered as best methods of coding in DNA based digital storage. A Huffman code translates binary to ternary digits and a rotating encoding translates ternary digits to nucleotides (Fig. 2). Huffman code that maps each binary byte to either 5 or 6 ternary digits. For

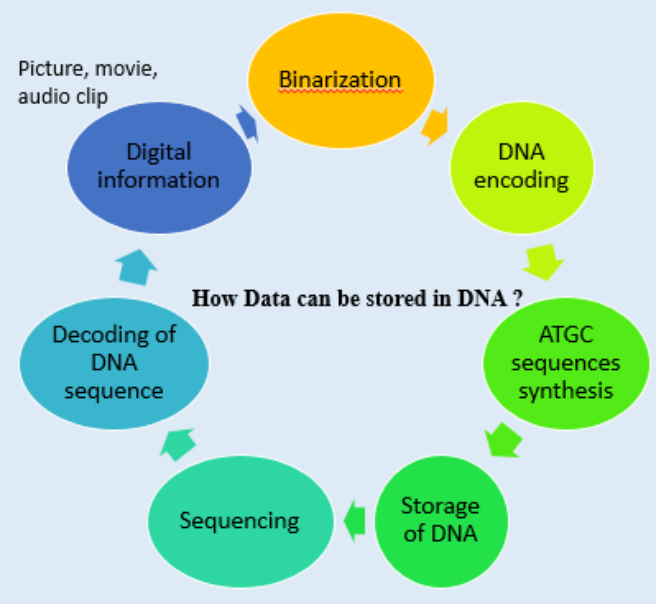

Fig. 1: Steps of DNA based digital information storage system

example, the Huffman code maps the binary string 01100001 to the base-3 string 01112. The rotating nucleotide encoding maps this string to the DNA sequence AGAGC.

Non-linear ternary error correction codes can be used instead of Huffman codes. The non-linear ternary codes described here improve the storage capacity of DNA by decreasing the length of the DNA required for storing a file. Once the digital file is encoded, the next step is to synthesize arbitrary single-strand DNA sequences chemically, nucleotide by nucleotide. There are several methods for *de novo* DNA sequence synthesis chemical oligonucleotide synthesis (phosphoramidite oligonucleotide synthesis) and oligo synthesis platforms using ink-jet printing (Agilent, Protogene), photosensitive 5′ deprotection (Nimblegen, Affymetrix), photolithography (Invitrogen) and electrochemical arrays (Oxamer) with 20 to 90 oligomers synthesizing capacity at least error rate (Tang *et al.,* 2013). Once we generate DNA oligomers or sequence that coded for digital information, DNA can be deep freeze at -20° to –80°C temperatures for a long time or till needed to decode through sequencing to recovered digital information.

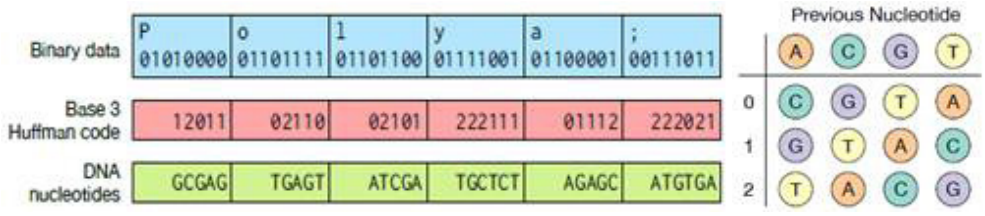

Fig. 2: A Huffman coding scheme

The goal of DNA sequencing is to read the DNA content, *i.e.,* to determine the exact nucleotides and their order in a DNA molecule. Sanger and Maxam-Gilbert sequencing technologies were first sequencing technologies used by biologists and still being used routinely in many laboratories. The emergence of second generation sequencing technologies *viz.* Roche's 454, Illumina, SOLiD, Ion Torrent and third generation sequencing technologies *viz.* PacBio and Oxford Nanopore opened new perspectives for genome analysis. However, in the past decade, the development of faster, cheaper, and higher-throughput sequencing technologies, "next-generation sequencing (NGS)" has dramatically expanded the reach of genomic studies. New sequencing technologies (next-generation sequencing) capable of producing sequences with very high throughput at a much lower cost about less than 0.1 USD for one MB of data and less than 1000 US Dollar for sequencing of the whole genome than the first sequencing technologies (Kchouk *et al*., 2017). The comparison of various sequencing technologies is presented in Table 3. Roche 454 and AB SOLiD systems are obsolete and currently not available in the market.

Oxford nanopore technologies have developed the world's first mini DNA sequencer MinION (Fig. 3). The MinION is portable in size as can carry just on the palm, real-time, generate long-read length, low-cost device. Oxford Nanopore readout electrical signals when a single-stranded DNA sequence passes through a nanopore made from proteins or synthetic materials resulting in to change in ion current which allow recognizing nucleotide bases. However, the error rate is significantly higher than other technologies (Yazdi *et al*., 2017).

Post-sequencing processing includes alignment of all reads using multiple sequence alignment (MSA) algorithms to obtain different consensus sequences using various bioinformatics tools, *viz*., Kalign, Clustal Omega, Coffee, MUSCLE, etc. (Yazdi *et al*., 2017). In the first phase of post-processing, the construction of a rough estimate of the DNA codewords using the address sequences. Initially, aligning all read results in a large number of errors but using different multiple sequence alignment (MSA) algorithms on the identified high-quality reads and obtained different consensus sequences. Each alignment method, new parameters were chosen by trial and error. The choice of the parameters was governed by the edit distance between the MSA consensus sequence and the corresponding DNA codeword. The consensus sequence is decoded as the reversal of the encoding process. The decoding process is simply the reverse of the encoding process. The procedure of post-processing, via sequence alignment and homopolymer correction, are presented in Figure 4.

Retrieval of DNA-based digital information storage

There are mainly following four approaches for retrieval of DNA based storage.

Church Approach

In this approach, user data was converted to a DNA sequence via a symbol-by-symbol mapping, encoding each data bit 0 into A or C, and each data bit 1 into T or G. Which of the two bases is used for encoding a particular bit is determined by a run length constraint, *i.e.,* one base is chosen randomly as long as it prohibits homopolymer runs of length greater than three.

Goldman *et al*. Approach

Goldman *et al*. approach started with a binary data set (Fig. 5). The binary file representation was obtained via

Table 3: The comparison of various sequencing technologies (Rhoads and Au, 2015)

	ABS	*Illumina*	*Ion Torrent*	*PacBio*	*Nanopore*
Instrument	DNA sequencer	MiSeq, Hiseq	PGM Ion Torrent	PacBio	MiniOn
Launch year	1987	2007	2010	2012	2016
Technology	Capillary electrophoresis	Sequencing by synthesis	Ion semiconductor	SMRT	Nanopore sequencing
No. of reactions	Detect 96 bases on time	25,000,000	10,00,000	50,000	10,00,000
Total data output/run	2.88 Mb bases /day	600 GB/run	300 Mb	400 Mb	500 Mb
Read length	900 bp	400 bp	200 bp	15000 bp	8,00,000 bp
Time/run	3–8 hours	3–10 days	2 hours	4–6 hours	2–3 days
Advantages	High quality, longer read	High throughput	Cheaper	High accuracy	Smallest size, portable
Disadvantages	High cost, low throughput	Short read size	Short read size	High cost	High error rate
Cost/million base	500 USD	0.04 USD	0.1 USD	0.8 USD	6–18 USD
Cost (INR)/base	0.0325	0.0000026	0.0000065	0.000052	0.00117

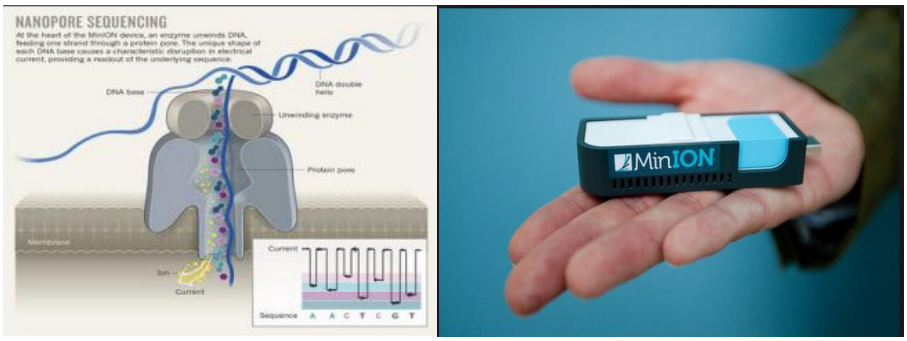

Fig. 3: Nanopore sequencing principle and Oxford's nanopore sequencer (MinION)

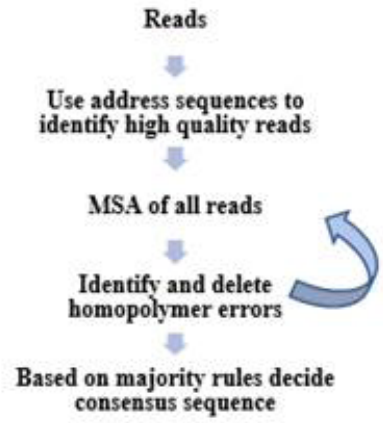

Fig. 4: Procedure of post-processing via alignment and homopolymer corrections

ASCII encoding, using one byte per symbol. Each byte was subsequently converted into 5 or 6 trits via an optimal Huffman code for the underlying distribution of the particular dataset used. Each trit is then used to select one out of three DNA oligonucleotides differing from the last encoded oligonucleotide. This form of differential coding ensures that there are no homopolymer runs of any length greater than one. Finally, the resulting DNA string is partitioned into segments of length 100 oligonucleotides, each of which has the property that it overlaps in 75 bases with each adjacent segment. This overlap ensures 4x coverage for each base. In addition, alternate segments of length 100 were reverse complemented. Indexing information, along with two trits for file identification, 12 trits for intra-file location information (which can be used to encode up to 314 unique segment locations), one parity-check and one additional base are appended to both ends to indicate whether the entire fragment was reverse complemented or not. The resulting fragment lengths of the constituent encoding amounted of length 117.

Grass *et al*. Approach

Grass addressed a specialized error-correcting scheme and best practices for DNA media maintenance. By this approach, we can store and recover information encoded in the DNA from the Global Seed Vault (at –18°C) even after 100 to 1000 years. They used error detecting and error correcting the Reed–Solomon encoding scheme. DNA encapsulated in silica appears to offer the most durable storage format, as silica has the lowest water concentration, and it separates DNA molecules from the environment through an inorganic layer. It also provides exceptional stability against oxidation and photo-resistance.

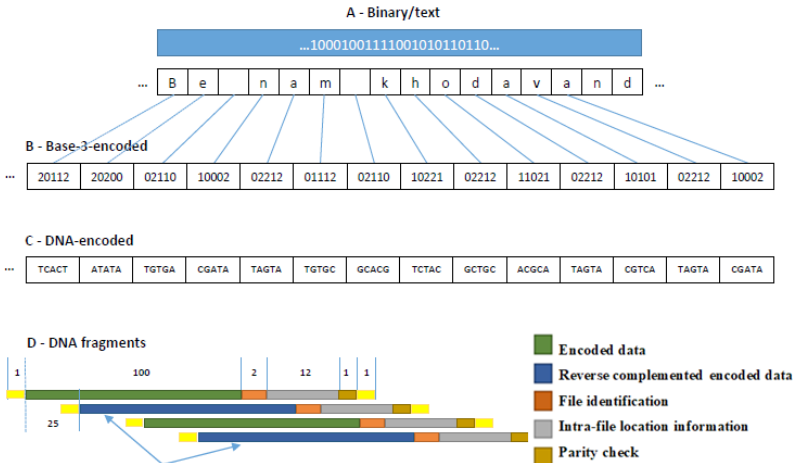

Fig. 5: Goldman *et al*. encoding method using ASCII and differential coding, Huffman compression with four-fold coverage

Yazdi *et al*. Approach of Random Access

Most prior DNA data storage efforts sequenced and decoded the entire amount of stored information, with no random access (Bornholt *et al*., 2016). However, this type of redundant sequencing becomes impractical as the amount of data increases. Bornholt *et al*. (2016), Yadzi *et al*. (2015, 2017) and Organick *et al*. (2018) successfully extracted selective files (e.g., retrieving only one image from a collection) of widely varying size and complexity from a DNA pool using PCR-based random access. Random access by mapping a key to a pair of PCR primers. At read time, those same primers are used in PCR to amplify only the strands with the desired keys which results into the pool will have a much higher concentration of the desired strands. Each chunk encoded with data addresses, payloads, and error detection codes and attaches the primer target sequences, to produce final DNA sequences for the synthesizer to manufacture (Fig. 6).

DNA INFORMATION DENSITY AND SHANNON INFORMATION DNA

Amount of data that can be encoded in DNA can be quantified by DNA information density. DNA information density measures the total amount of data that can be stored in a unit gram of DNA. At theoretical maximum, one gram of single-stranded genetic code can encode 455 EB (10^{18}) of information. Goldman *et al.* (2013) stored 2.2 PB (10^{15}) of digital information per gram of DNA, while Limbachiya *et al*. (2016) stored about 1.15 x 10^{20} = 115 EB (10^{18}) of digital information using DNA Golay codes. Following is the formula for calculation of DNA information density.

Shannon information for DNA is a number of bits encoded per DNA base; if higher data density, higher DNA Shannon information, which results in lower DNA storage cost. DNA Shannon information = Size of files (bytes)/(number of chunks x Chunk size). For given file of size 757051 bytes, number of chunks is 84126 and chunk size is 112 Shannon information = (757051) / (84126 x 112) = 0.081 bytes/base. The majority of costs are the cost of DNA sequence synthesis and sequencing, which limits DNA as a digital information storage medium. Different coding scheme affect the cost of DNA based storage system i.e., DNA Golay code required less amount of the DNA which reduce the cost of this technology (Limbachiya *et al.,* 2016) and using this coding method, negligible increase in cost per unit data storage with the increase in the amount of data compared to Goldman approach. Various scientists reported different costs of DNA based digital information storage (Table 4).

CASE STUDIES OF DNA BASED DIGITAL INFORMATION STORAGE

Various scientists have used different digital information in the form of audio clips, images, pdf files, video files, old documents, text files, etc. and encoding scheme for DNA based digital information storage. The schemes are presented chronologically based on the publication date in Table 5. Different scientists used different size of information ranges from 0.003 to 200 MB. They used different encoding schemes, *viz.,* One bit one base, Alternating code, Huffman's code, Golay codes and Reed Solomon coding schemes to encode digital information. The sequencing technologies, sequencing coverage, error corrections and detections, DNA information density, and whether the approach was random access or not, etc. are presented in Table 5.

CHALLENGES OF DNA BASED DIGITAL INFORMATION STORAGE

DNA would become a universal storage medium in the future, but presently it has several challenges because of its own physical as well as chemical composition or because of technological incompetence. There are several challenges of DNA based digital information storage, which are presented below (Shrivastava and Baldani, 2014):

- The overall process of encoding, amplifying, sequencing, and decoding takes significantly more time.

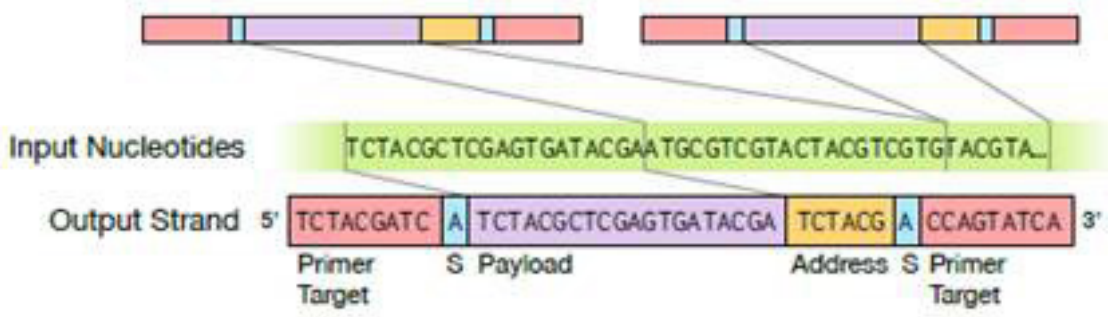

Fig. 6: Yazdi *et al*. approach of random access

Table 4: Costs of DNA based digital information storage

File size	*Cost*	*Cost/byte*	*Cost/bit in USD*	*Cost/byte (INR)*	*Author*
15 MB	0.2 million USD	0.127	0.016	8.3	Shah *et al*., 2014 (DAIICT)
150 MB	1.8 million USD	0.012	0.0015	0.78	
1 MB	12,620 USD	0.012	0.0015	0.78	Goldman *et al*., 2013
1 MB	3500 USD	0.0033	0.00041	0.22	Erlich and Zielinski, 2016
200 MB	60 million USD	0.29	0.036	18.85	Microsoft, 2015
12 MB	0.1 million USD	0.080	0.010	5.2	Twist Bioscience

Table 5: Comparison of DNA storage coding schemes and experimental results

Scientist	*File size (MB)*	*Encoding scheme*	*Sequencing technology*	*Fold coverage (seq. depth)*	*Sequence error rate*	*Error detection*	*Bits/ base*	*Random access*
Church *et al*., 2012	0.65	one bit per base	HiSeq	3000	0.1–0.3%	None	0.83	No
Goldman *et al*., 2013	0.76	Huffman code	HiSeq	51	0.1 %	Detection	0.33	No
Yazdi *et al*., 2015	0.003	Alternating code	Sanger's	200	0.05 %	Correction	1.575	Yes
Limbachiya *et al*., 2016	0.76	Golay code	–	–	–	None	0.73	No
Grass *et al*., 2015	0.08	Reed Solomon	MiSeq	372	0.1%	Correction	1.16	No
Bornholt *et al*., 2016	0.15	Huffman code	MiSeq	40	0.1%	None	0.85	Yes
Erlich and Zielinski, 2016	2.11	–	MiSeq	10.5	0.1%	None	1.55	No
Yazdi *et al*., 2017	200	–	MinION		12%	Correction	1.74	Yes
Organick *et al*., 2018	200	Reed Solomon	Illumina	5	-	-	1.10	Yes

- Retrieval process much slower than that of a personal computer (which can read data from the hard drive at nearly 100 Mb/second), consequently, DNA is unlikely to compete with optical, magnetic-based storage media.
- Many types of errors are associated with the current DNA synthesis and sequencing technologies, viz., homopolymers, sequencing errors, error due to lower access rate, etc. DNA in living cells have auto-correction enzymes, and no such artificial enzymes exist for artificial DNA.
- DNA Strings need to be discarded if the decoding scheme is inefficient, thus leading to a loss of data and consumption of more DNA to ensure the same theoretical completeness.
- Due to DNA's structure, it is prone to mutations in extreme conditions; thus the data might get altered in a mutation.
- Another major challenge for practical DNA-based information storage is the difficulty of synthesizing long sequences of DNA de novo (simulating on the computer) to a specified design.
- Very high cost of DNA synthesis and retrieval of information.
- Sequences can also degrade while stored, so data integrity is questionable.

Future prospects

In the future, with an efficient encoding, error corrections, cheaper DNA synthesis, and sequencing, DNA based storage will become a practical solution for storage of exponentially growing digital data. For making DNA based digital information storage practically feasible, it needs to develop cheaper DNA synthesis technology, to increase the data storage capacity of DNA and to make biochemical and computational improvements to make it possible to read/write EB of data within hours or even days.

References

Beck, M.B. and Yampolskiy, R.V. (2015). Hiding color images in DNA sequences. Paper presented at the 26th modern artificial intelligence and cognitive science conference (MAICS 2015), Greensboro, North Carolina, 118-120.

Bioscience: DNA data storage by Bioscience. Retrieved from http://www.twistbioscience.com /products/storage.

Bornholt, J., Lopez, R., Carmean, D.M., Ceze, L., Seelig, G. and Strauss, K. (2016). A DNA-based archival storage system. Proceeding of International Conference ASPLOS: 637–649. http://dx.doi.org/10.1145/2872362.2872397.

Church, G.M., Gao, Y. and Kosuri, S. (2012). Next-generation digital information storage in DNA. Science, 337 (6102): 1628.

Erlich, Y. and Zielinski, D. (2016). DNA fountain enables a robust and efficient storage architecture. Science, 355: 950-954.

Goldman, N., Bertone, P., Chen, S., Dessimoz, C., LeProust, E.M., Sipos, B. and Birney, E. (2013). Towards practical, high-capacity, low maintenance information storage in synthesized DNA. Nature, 494 (7435): 77-80. doi:10.1038/nature11875.

Grass, R.N., Heckel, R., Puddu, M., Paunescu, D. and Stark, W.J. (2015). Robust chemical preservation of digital information on DNA in silica with error correcting codes. Angewandte Chemie International Edition, 54 (8): 2552–2555.

Kchouk, M., Gibrat, J.F. and Elloumi, M. (2017). Generations of sequencing technologies: From first to next generation. Biol. Med., 9: 395. doi:10.4172/0974-8369.1000395.

Laddha, R. and Honwadkar, K. (2016). Digital data storage on DNA. Intl J.Computer Applic., 142 (2): 43-46.

Limbachiya, D. and Gupta, M.K. (2015). Natural data storage: A review on sending information from now to then via Nature. ACM Journal on Emerging Technologies in Computing Systems, 5: A1-A17.

Limbachiya, D., Dhameliya, V., Khakhar, M. and Gupta, M.K. (2016). On optimal family of codes for archival DNA storage. Proceeding of Seventh International Workshop on Signal Design and its Applications in Communications (IWSDA), 1-6. (doi: 10.1109/IWSDA.2015.7458386).

Microsoft: DNA storage by Microsoft (2015). Retrieved from https://www.microsoft.com/en-us/research/project/dna-storage/.

Organick, L., Ang, S.D., Chen, Y.J. Ceze, L. and Strauss, K. (2018). Random access in large-scale DNA data storage. Nature Biotechnology, 1-13. doi:10.1038/nbt.4079.

Rhoads, A., & Au, K.F. (2015). PacBio Sequencing and Its Applications. Genomics Proteomics Bioinformatics, 13, 278–289.

Sandisk technology (2014). Currently available memory-cards as storage devices. Retrived from www.sandisk.com/home/memory-card)www.sandisk.com/home/memory-card.

Sony Develops Magnetic Tape Storage Technology with the Industry's Highest Recording Areal Density of 201 Gb/in2. retrieved from https://www.sony.net/ SonyInfo/News/Press/201708/17-070E/index.html.

Shah, S., Limbachiya, D. and Gupta, M.K. (2014). DNACloud: A tool for storing big data on DNA. Proceeding of 11th Annual Conference on Foundations of Nanoscience: Self-Assembled Architectures and Devices, 1-6.

Shrivastava, S. and Badlani, R. (2014). Data storage in DNA. Intl J. Electrical Energy, 2(2): 119-125. doi: 10.12720/ijoee.2.2.119-124.

Tang, N., Ma, S. and Tian, J. (2013). New tools for cost-effective DNA synthesis. Synthetic Biology, 3-18. doi: http://dx.doi.org/10.1016/B978-0-12-394430-6.00001-7.

Yazdi, S.M.H.T., Yuan, Y., Ma, J., Zhao, H. and Milenkovic, O. (2015). A rewritable, random access DNA-based storage system. Science Report, 5: 14138.

Yazdi, S.M.H.T., Gabrys, R. and Milenkovic, O. (2017). Portable and error-free DNA-based data storage. Scientific Reports, 7 (5011): 1-6. doi:10.1038/s41598-017-05188-1.

Yim, A., Yu, J., Li, A., Wong, J., Loo, K., Chan, S., Kong, K., Yip, A. and Chan, T. (2014). The essential component in DNA-based information storage system: Robust error-tolerating module. Frontiers in bioengineering and biotechnology, 2: 49.