Journal of Forensic Sciences And Criminal Investigation
ISSN: 2476-1311

Juniper Publishers key to the Researchers

Review Article
Volume 15 Issue 3 - September 2021
DOI: 10.19080/JFSCI.2021.15.555914

J Forensic Sci & Criminal Inves

Storing Photos and Other Documents in DNA And Its Use in Forensic Science - A Review

Ritoza Das*

Department of Forensic Science, Jain deemed-to-be University, India

Submission: September 04, 2021; **Published:** September 22, 2021

***Corresponding author:** Ritoza Das, Department of Forensic Science, Jain (deemed-to-be) University, Bangalore - 560027, Karnataka, India

Abstract

DNA or Deoxyribonucleic acid is a molecule that is present in almost all living organisms and forms the basis of the development and functioning of an organism. It is a huge molecule with millions of atoms and contains genetic material. One of its main functions is storing information. It contains long chains of nucleotides – A, T, C and G. Data can be stored in the sequence of these letters, turning DNA into a new form of information technology. Along with this, it has a large storage capacity. With the amount of data that is being produced per day, it is becoming harder to find a cheaper way to store this massive amount of information. Therefore, scientists have found an alternative way to combat this problem. According to researchers in Harvard University, DNA can store 455 exabytes per gram (1 exabyte = 1018 bytes). Subsequent research on this matter can alleviate the issue of storing the ever-growing amounts of data that is being generated. Since DNA is stable and can store data for a long time, it can be used to store all types of data without the fear of it being deleted. If this can indeed be achieved, it would make forensic investigations easier than it already is. Instead of getting matching sequences, we could get to see inside the life of a perpetrator/suspect if they leave their DNA behind in the form of blood, hair, skin tissues etc. The DNA extracted from these can be compared with the synthetic DNA of the respective person where all their data is stored. Photos, videos, and other documents if stored, located, and extracted properly in the DNA can be of significant use for maintaining the storage of excess data while also aiding officials in tracking down criminals and missing persons. Even from a cybersecurity point of view, this can be beneficial as DNA steganography can be performed by hackers. This paper shows how much of this can be achieved with today's technology and its growing importance in the future.

Keywords: DNA; Data; Steganography; Storage; Information; CRISPR; Forensic science

Introduction

DNA is essential for all living organisms. It contains information that helps us develop, survive, and reproduce. The main functions of DNA include replication, mutation/recombination, gene expression and encoding information. It is sometimes deemed as a blueprint, since it contains the instructions to construct other components of the cell, such as proteins and RNA molecules. But now instead of just storing information related to vital life processes, it can be used to store all our personal information like pictures, videos, songs, bank statements etc. Presently, Earth has more than 10 trillion bytes of data and this keeps on increasing day by day. We are generating an average of 2.5 million gigabytes of data per day. These copious amounts of data are usually stored in Exabyte data centers which cost a lot ($ 1 billion) to maintain and are usually the size of football fields (hence, taking up a lot of space). International Data Corporation has forecasted that the global data storage demand will grow 175 trillion gigabytes by the year 2025 which will exceed the storage capacity of currently available storage devices, such as hard drives, pen drives, and optical discs. Magnetic tapes which form the basis of most digital archives are approaching their density limits, have extremely limited life spans and can be damaged easily. In this hour of need where we are running out of spaces and spending a generous sum of money to maintain and store all this information, an alternative is needed [1]. Here, DNA comes into the picture. The reason scientists have been curious about DNA as a potential new storage medium other than machine memory is because of its multiple advantages. DNA has ultra-high-density storage, it is very stable (due to its double polymer structure) and we can store data for a prolonged period (even for a millennium in some cases). In 1959, Richard Feynman talked about synthetic DNA as a promising candidate and the creation of artificial objects like biological objects performing the same functions. Computers and organic

cells have been seen to have a lot in common. In computers, information is encoded in the form of binary digits (1s and 0s) and these help in executing programs. In cells, the information is encoded in the form of four nucleotide bases – Adenine (A), Thymine (T), Cytosine (C), and Guanine (G). When read, these help in the production of proteins. DNA storage involves 2 main processes.

I. DNA synthesis (writing the code)

II. DNA sequencing (reading the code)

If effectively used, the digital information which is now being stored in warehouse-sized data centers could be stored in a space almost similar to the size of a tiny sugar cube and will take very little energy to maintain. The human genome contains 3 billion bp of DNA. This is similar to carrying 1.5 GB in each diploid human cell. As for its use in forensic science, it could provide a lot of information related to an individual in a very less amount of time – something which can be very less time consuming and could produce results extremely fast Figure 1.

Type	Life Expectancy	Capacity
DNA	Millions of years	10^8 TB per 1 gram [1]
Hard disk	~10 years	Up to 4 TB (2011)
CD	~10 years	800 MB
DVD	<10 years	Up to 17GB
USB flash drive	~10 years, depending on usage	Up to 256GB (2011)
Tape	~30 years	Up to 35 TB

Figure 1: Table representing the life expectancy and storage capacity of various storage media.

Review of Literature

1) Reinhard Heckel, Gediminas Mikutis & Robert N. Grass, 2019 talk about the longevity and enormous density of DNA and how this makes it a promising archival storage medium. Although due to some constraints, the data can only be written in many short DNA molecules which are to be stored unorderly. Also, imperfections in writing, reading, storage and handling of DNA may lead to loss of DNA molecules. Hence, to store DNA in effective DNA storage systems, a qualitative and quantitative understanding of the errors is very crucial.

2) Luis Ceze, Jeff Nivala and Karin Strauss, 2019 survey the field of in-vivo molecular memory systems that help in recording and storing information in the DNA of living cells. Along with this, they anticipate technological innovations that are tailored for DNA data storage, which promise to gradually decrease barriers to its mainstream adoption.

3) Marc B Beck, Eric C Rouchka, Roman V Yampolskiy, 2012 in their research talk about various methods used to insert information in DNA sequence for storing data, watermarking, or communicating messages secretly. Along with this, they developed a software toolkit to determine hidden messages in DNA sequences.

Methodology

The process of storing data in DNA

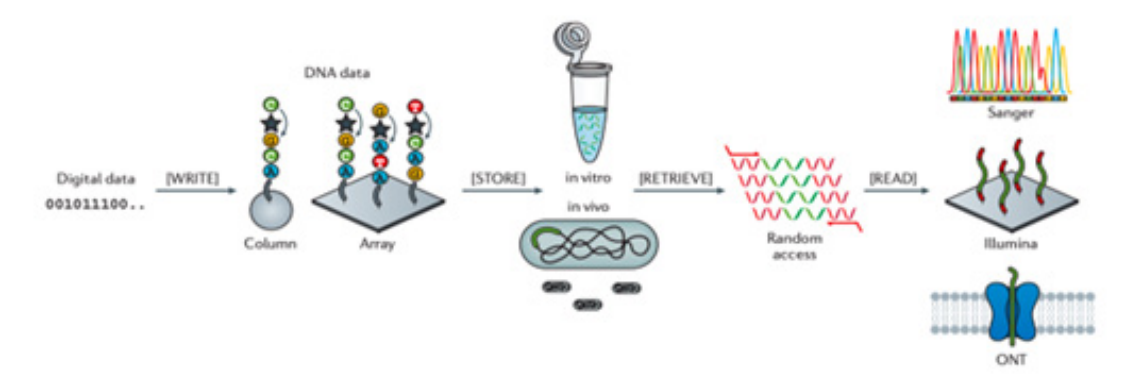

Figure 2: Diagram representing the encoding and decoding process of data in DNA [11].

How to cite this article: Ritoza D. Storing Photos and Other Documents in DNA And Its Use in Forensic Science - A Review. J Forensic Sci & Criminal Inves. 2021; 15(3): 555914. DOI: 10.19080/JFSCI.2021.15.555914

DNA storage mainly contains 2 processes. They are DNA synthesis and DNA sequencing. The binary code from a computer is first translated to four base pairings (encoding) and then the DNA molecules are synthesized letter by letter by using enzymes or with the help of chemical reactions. These are at last indexed from where they can be selectively accessed (random access). After this, they are stored in containers in which temperature and light conditions are regulated to maintain stability. DNA is usually stored by freezing it in a solution, by drying or encapsulating in a bead. Later, a DNA strand is selected and decoded by a commercial sequencing machine (which was initially developed for genome sequencing) and then translated back to the original file (decoding). Error collecting algorithms are used so that the data recovered is as error-free as possible Figure 2.

Advantages of DNA as a storage medium

a) Listed are the various advantages of DNA as a storage medium -

b) It is way denser than other forms of memory up to 1018 bytes per mm3.

c) It is easy to synthesize and sequence.

d) We can store data for a long time (if proper protocols are followed, then a millennium).

e) It can't be destroyed easily.

f) Multiple copies of the data can be generated (ease of replication).

Disadvantages of DNA as a storage medium

Listed are the disadvantages of DNA as a storage medium

a. Expensive to initiate (although later the costs are comparably lower.)

b. It is difficult to pick up data and view it.

c. Susceptible to high rate of errors.

d. Using PCR (Polymerase Chain Reaction) can lead to "cross-talk" i.e., unwanted files can be pulled out.

Retrieving Information from DNA

In the current scenario, DNA is retrieved using Polymerase Chain Reaction (PCR) since each DNA data file contains a particular sequence that gets bound to a particular PCR primer. To pull a particular file, primer is added to the sample. This helps to find and amplify the desired sequence. One drawback is that this can lead to crosstalk between primer and off-target sequences, hence unwanted files are pulled out. PCR also requires using enzymes and thus consumes most of the DNA from the gene pool [2]. An alternative approach to this was performed by scientists from MIT, where they encapsulated each DNA file into small silica particles. Each capsule is labelled with single-stranded DNA "barcodes" that correspond to a single file. The capsules can fit information up to 1 gigabyte. Each barcode corresponds to labels like "bat" or "cup". When researchers want to pull out a specific image, they just need to remove a sample of DNA and add the primers which correspond to the labels wanted. Example – "cat", "yellow" and "wildlife" may be used to extract a picture of a tiger. If we put two of these labels in a file, we can uniquely identify and label 10 billion different files individually and if we put four of these labels, then we can uniquely identify 1020 files. The primers are also labelled with fluorescent/magnetic particles which makes it convenient to pull out and identify any matches from the sample itself. The retrieval process also allows Boolean logic statements to generate results. The file system's search rate is determined by the data size per capsule and is limited as of now due to the cost to write even a minimum of 100 megabytes of data. This kind of DNA encapsulation can help in storing archived data or "cold" (something which is not needed that often) [3].

Crispr Gene Editing

Clustered Regularly Interspaced Short Palindromic Repeats or CRISPR Gene Editing is a genetic engineering technique used commonly in molecular biology. It has 2 components.

i. Cas9 nuclease - a DNA cutting protein

ii. Guide RNA - RNA molecule

These two bounds together to form the Cas9 complex which identifies and cuts specific sections of DNA. It locates and binds to a common sequence in the genome called PAM. After the binding, the guide RNA unwinds a part of the double helix of DNA (the RNA is designed to match and bind particular DNA sequences). After finding the correct sequence, Cas9 cuts the DNA i.e., its nuclease domains make a double-stranded break. The cells try to fix this break, but this is error-prone and often induces mutations leading to the disabling of the gene itself. Hence, CRISPR is a great tool for cutting out specific genes. Sometimes it is used by attaching fluorescent proteins to the Cas9 complex to see where the DNA sequences are found in the cell. This can help in the accurate cutting of the DNA sequences which we prefer to pick out [4] (Figure 3). This technique can also be used in in-vivo recording and data storage systems. Cutting and repairing of target sites can result in unique changes (point mutations), insertions and deletions that eventually serve as "barcodes". This complex also records a non-binary range of mutations or barcodes into the DNA over time. Another complex Cas1-Cas2 complex, which is an integrase and responsible for integrating small viral DNA pieces into precise locations in the cellular genome can be used too. The digital data is encoded into pools of short synthetic DNA segments and introduced into cells expressing Cas1-Cas2. The synthetic DNA sequences are integrated into the CRISPR array within the cellular genome. This can be repeated over time, introduce unique data during each round. This approach was the first definitive case of an in-vivo recording method that enabled encoding and decoding of digital data. Arbitrary data can be stored in defined sequences and uploaded in the genome on demand. Digital data

How to cite this article: Ritoza D. Storing Photos and Other Documents in DNA And Its Use in Forensic Science - A Review. J Forensic Sci & Criminal Inves. 2021; 15(3): 555914. DOI: 10.19080/JFSCI.2021.15.555914

is often encoded within synthetic oligonucleotides and integrated within the CRISPR array. Even though a lot is left to achieve practically, this method can be used to store information in DNA while also making it easier to access, inside human beings or other live organisms.

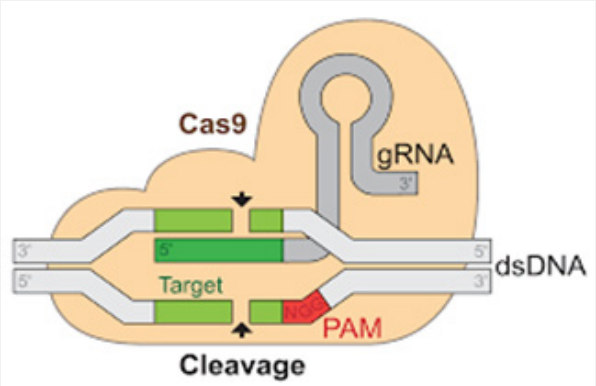

Figure 3: CRISPR Cas9 Gene Editing process.

Significance In Forensic Science

Studies on DNA data storage systems still have a lot left to be found and applied. But it does have a lot of potential in making investigations easier for forensic scientists. Data stored in DNA in living organisms can help get a look into the life of a possible perpetrator or victim. There are many ways to implement this to aid the judicial system in the coming years. It has already been seen that living organisms are a great source of storing information [5]. In 2007 showed how it's possible use DNA of living organisms as a storage medium by encoding the message "E=mc^2 1905!" in the genome of B. subtilis. A database could be constructed to store each person's data according to their unique genetic blueprint containing all their own data. This can be a storage format itself to store all our data in DNA sequences and would be easy to find out for everyone. Biological samples found in the crime scene like - blood, saliva, semen, skin tissues, urine etc. provide us with the DNA sample of an individual. Instead of trying to get a match, a scientist can just take the DNA sequence and match it with the sequences on the database. This will give us an inside view of the life of that person. Based on just a tiny bit of sample, it can help identify.

a. Who it belongs to

b. Their family

c. Locations they've been

d. Documents they have saved (cheques, medical reports etc.)

e. Songs/Videos

f. People we know/met

g. Books they've read

This is more than a GED match using genetic genealogy can offer. Using this application, a person can be directly singled out instead of going through tons of forensic examinations. A look inside the perpetrator's/victim's life can provide valuable information regarding their lifestyle and the case which can save a lot of time in criminal investigations. There are also ways in which criminals can take advantage of this. This is being done through a technique called - Steganography. It is the method through which secret messages are hidden. While cryptography makes the message unreadable, steganography completely hides the message so as to not raise suspicion. Hackers and spies can send confidential information stored in DNA and send it out in an unsuspected biological sample, all with proper information and access to technology. Not only images but other forms of digital data (including malicious codes) can be sent this way. The non-coding region is a popular choice to store this data. Codons are often used to embed the messages. Although, using steganalysis can help in weeding out the hidden messages. A software known as "Word Spy" was discovered by Wang and Zhang which can detect the biological features in a genome as secret messages. Another software called "DNA - Steg" has 2 programs used for steganography and steganalysis while also offering different coding schemes. Similarly, other software's can also insert and extract messages from DNA sequences and some can detect modified versions of the DNA sequence with the original. These can be used to develop new methods to detect steganography. Further research is still ongoing.

Limitations And Discussion

Of course, the use of DNA data storage in daily life and forensic science is only possible if in-vivo or synthetic DNA data storage systems become more efficient and a practical method of storage in the future. Along with this, another obstacle which DNA data

How to cite this article: Ritoza D. Storing Photos and Other Documents in DNA And Its Use in Forensic Science - A Review. J Forensic Sci & Criminal Inves. 2021; 15(3): 555914. DOI: 10.19080/JFSCI.2021.15.555914

storage might face is its practicality. The common man has yet to feel the need of storing their private data into their DNA as it cannot be accessed by just anyone. It would require professional help. But the greatest difficulty this method will face is ethicality. Even now, some people are divided between their views towards law officials using GED matches in catching criminals, some viewing it as an exploitation of their privacy, so in order to store their own data in DNA and keep it in a database would be something which will be straight outrageous. Also, effective storage of larger amounts of data like the data produced by a whole city or state is yet to be achieved. But with the growing importance of finding better and cheaper storage options and a technologically paced world, all this can soon be achieved and hopefully deemed useful by everyone [6-11].

Conclusion

In this review paper, we have talked about the advantages and disadvantages of using DNA as a storage medium and how it can help solve the ever-growing data storage problems in the world. The stability of DNA and the ability to store data for a longer time than the traditional storage devices have made it an important commodity for scientists. Many methods and techniques like PCR, CRISPR etc. can be used in encoding, decoding and retrieval of information from DNA sequences and how it can be used by forensic scientists to get closer to justice and hackers in their professions. Although the full extent of its working is yet to be achieved, DNA data storage systems have a bright future if thought out and executed properly.

References

1. De Ridder CA, Morton JD (2020) When Will DNA Solve. The Data Storage Crisis? Pillsbury
2. Trafton A (2021) Could all your digital photos be stored as DNA? MIT News.
3. (2020) Finding Data in DNA: Computer Forensic Investigations of Living Organisms.
4. Ran FA, Hsu PD (2020) Genome engineering using the CRISPR-Cas9 system.
5. Yachie N, Sekiyama K (2020) Alignment-based approach for durable data storage into living organisms.
6. Grigoryev Y (2012) How Much Information is Stored in the Human Genome.
7. (2020) What are genome editing and CRISPR-Cas9?
8. Dong Y, Sun F, Ping Z (2020) DNA storage: research landscape and future prospects. National Science Review.
9. Akram F, Haq IU, Ali, H (2018) Trends to store digital data in DNA: an overview.
10. Silva PYD, Ganegoja GU (2016) New Trends of Digital Data Storage in DNA. BioMed Research International.
11. Ceze L, Nivala J (2019) Molecular digital data storage using DNA. Nature Review Genetics.

DOI: 10.19080/JFSCI.2021.15.555914

How to cite this article: Ritoza D. Storing Photos and Other Documents in DNA And Its Use in Forensic Science - A Review. J Forensic Sci & Criminal Inves. 2021; 15(3): 555914. DOI: 10.19080/JFSCI.2021.15.555914