

# Preventing Data Loss by Storing Information in Bacterial DNA

**Mohan S**  
PG scholar, PSG College of  
Technology  
Department of Information  
Technology  
Coimbatore 641004

**Vinodh S**  
PG scholar, PSG College of  
Technology  
Department of Electrical and  
Electronic Engineering  
Coimbatore 641004

**Jeevan F R**  
PG scholar, PSG College of  
Technology  
Department of Information  
Technology  
Coimbatore 641004

## ABSTRACT

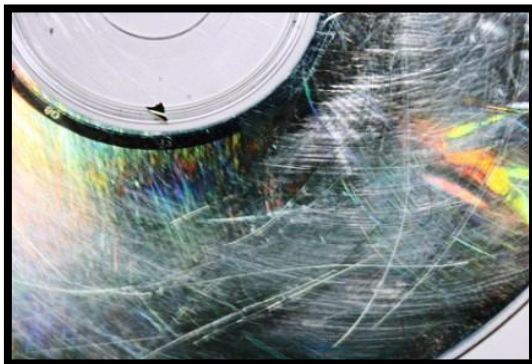
One of the major problems faced today in data storage is the data loss due to crashing of storage devices like magnetic disc or optical disc. In recent years scientists have turned their attention towards the biomaterials for data storage. Attempts were made to store data in proteins, tissues, etc. But the problem faced over these things are, that they are not reliable, easily biodegradable and information cannot be transmitted over generations. This paper discusses about a new technology of storing data in bacterial DNA. This is achieved by the process called induced mutation which is controlled by electrical signals. Data stored in DNA is more reliable than any other devices. This invention provides a hope for future data storage which will be safe for millions of years, as they are transmitted over various generations. We can cultivate the bacteria in which the data are stored in various places like fish tank, garden, stomach, etc.

## Keywords

Luminometer, Chelex's process, Mutation, Bioluminescence, Intercalating Agent

## 1. INTRODUCTION

One of the major researches in data storage is done in BioStorage Technologies. It refers to encoding digital information on a biological molecule. It is half 'Information Technology' oriented and the rest 'Biotechnology'. Earlier, we have discovered storing data on protein, blood tissue, etc. It is not reliable as the biological material can degrade over time. In order to overcome this defect, we store data in DNA, by which data can be transmitted from one generation to another.[9] So we go for storing data on genetic material which can be transmitted for several generations.



**Fig: 1 Damaged Optical Disk (CD)**

DNA (Deoxyribo Nucleic Acid) is a very good content that contains lots of genetic material and good natural biological data storage medium. It contains all the information about an organism. The organism's complete details like the physical structure, ancestral information, capability, etc. are naturally stored in this type of materials. Any change in it can change the complete nature of the organism. But some part of DNA is not used throughout its life. Such portion of DNA is identified and data is stored in it. Now we are choosing bacteria DNA for data storage. We can discuss about this deeply further.



**Fig: 2 Magnetic Disc (Hard Drive) Damaged**

## 2. LITERATURE SURVEY



**Fig: 3A Typical Bacterium that can be used**

Researchers from the Japanese universities, Keio University Institute for Advanced Biosciences and Keio University Shonan Fujisawa Campus in Tokyo have succeeded in creating an artificial DNA containing the data to be preserved. They published their invention on 27 Feb 2007.[5]

The study, published in the academic journal *Biotechnology Progress*, describes how DNA encoded with "E=MC2 1905!" was put into the genome of the common soil bacterium *Bacillus subtilis*. While ink may fade and computers may crash, bacterial information lasts as long as a species stays alive — possibly a mind-boggling million years — according to Professor Masaru Tomita, who heads the team of researchers at Keio University.

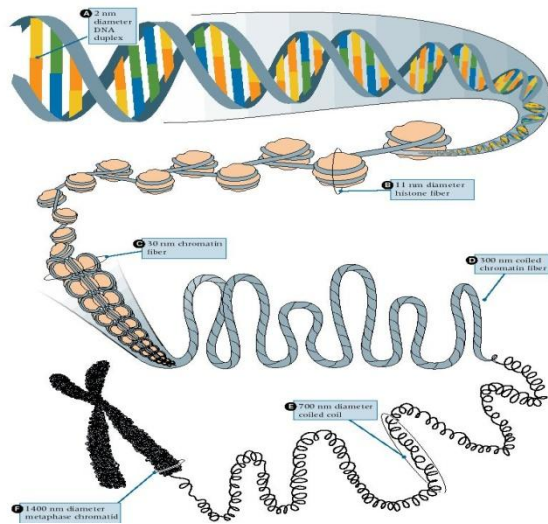


Fig: 4 DNA Structure

### 3. DNA

DNA storage is hierarchical and multilayer.[9] We may say it consists of files grouped into folders which are organized into volumes of data. Every animal cell has a nucleus (the central and controlling part of the cell). This contains some fixed amount of chromosomes (depending upon the organism). These chromosomes have many genes. Each gene is made up of millions of DNA. These nucleic acids are just complex organic molecules. As per Watson and Crick (persons who designed the DNA structure at first) it is a helical ribbon structure. There are two ribbons on either side, each contains alternatively placed sugar and protein molecules. These proteins are bonded with Adenine, Guanine, Thiamine and Cytosine.) Any small change in the environment can easily affect this molecule which may sometimes undergo mutation. Mutation is a small or may be even bigger sudden change in the genome sequence. This may even change the entire organism. This mutation helps us in data storage.

The DNA is as stable as its living carrier (bacteria, plant or animal cell). In some cases after the cell's death the DNA data may survive. In suitable conditions the DNA may hold its data for years. On the other hand the DNA is vulnerable to hydrolysis and oxidation. Many factors lead to mutations in the genes. In mutations, letters of the genetic code can be changed and stretches of DNA can be deleted. The mutation rate of DNA replication process of *Escherichia coli* is near  $10^{-9}$ . Some biological cells are more resistant to environmental conditions than others. For example bacteria such as *Deinococcus radiodurans* are especially good at surviving extreme conditions. They can tolerate high temperatures, desiccation and ultraviolet and ionizing radiation doses about 1000 times higher than fatal ones to humans. The DNA data protection mechanism has two basic layers. The genetic code

is to some extent redundant and the genetic information is massively replicated. Most of the amino acids are encoded by more than one codon (in total 61 codons are used to code 20 amino acids), for example alanine is coded by four different codons. On the other hand there are amino acids (such as methionine) which are coded by single codons. Furthermore, the same DNA information is replicated in every cell of the particular living organism.

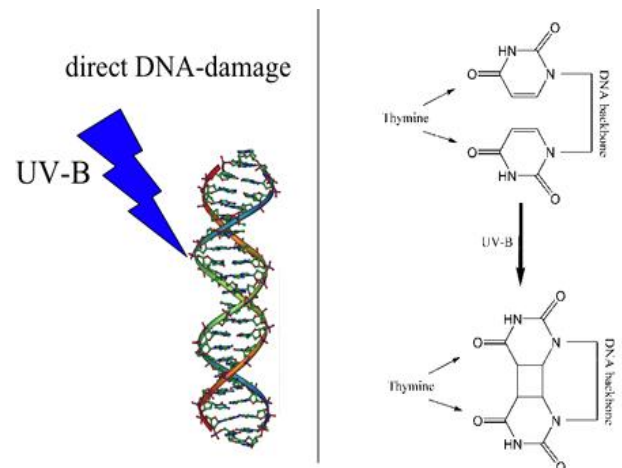


Fig: 5 DNA Responding to UV rays

### 3.1 Mutation

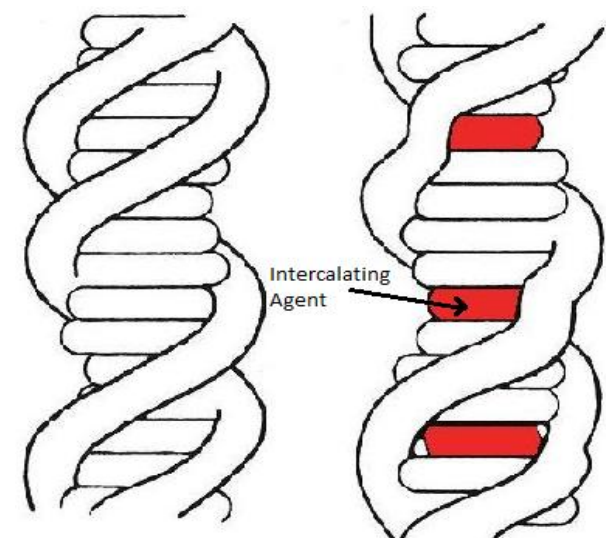


Fig: 6 Formation of Intercalating Agent in DNA

Mutations are heritable changes in genetic material that may be spontaneous or induced. In the present study, we make the DNA to mute by radiations, which is induced mutation. Ultra violet rays are ionizing radiations to mute the DNA. We encode the data into the genome. This small change in the DNA can be transmitted over generations so that the data will live along with the molecule. This mutation results in evolution.

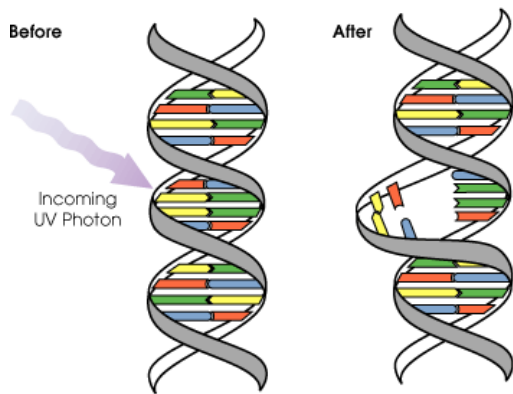
### 3.2 Mutation by Ultraviolet rays

When the ultraviolet rays fall on the DNA it absorbs some energy. This energy excites the electrons in the DNA to a higher energy level. As a result it results in mutation. The two nearby thiamine molecules get bonded with each other leaving the adenine molecule. This forms an intercalating agent in the

DNA, which is unstable. This instability causes the DNA to get muted. Again when it gets into the normal state the genome sequence is changed and results in mutation. This is induced by ultraviolet rays. The role of ultraviolet rays is to increase the mutation rate. Now here we control the mutation rate by digitally made electrical signals. This forms the basic principle in data storage in DNA.

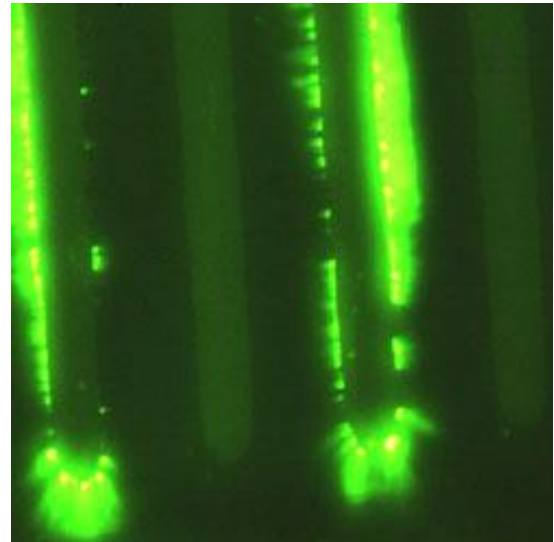
#### 4. DATA ENCODING IN DNA

At first a suitable bacterium is chosen which must be chosen in a way that it must tolerate high temperatures, desiccation, ultraviolet light and ionizing radiation doses 1000 times higher than would be fatal to humans.[2] Bacteria like *Escherchia coli*, *Deinococcusradiodurans*, *Serratiamarcescens*, *Bacillus subtilis*, etc are normally used as they can tolerate extreme conditions. The DNA molecule is extracted from it by Chelex's process (A well known biotechnological method to collect bacterial DNA on water on a test tube). Then it is stored in a proper temperature. Normally 20oC is maintained. Then the DNA is cloned (taken synthetic copies of its own by biotechnological method). The synthetic DNA is taken for processing of data storage.



**Fig 7: DNA in data acceptance form**

Then the DNA is taken for Bioluminescence.[3] It is a process of taking the DNA to a state by which it is ready to accept electrical signals. The process of Bioluminescence contains the steps that are as follows.[4] The DNA is first heated to 30oC in a biological medium. Then the DNA is brought to luminescence state. In the luminescence state the data will be ready to receive electrical impulses.[5] This is done by using some chemical substances that makes the DNA to a luminescence state. Normally aldehyde n-decanal is used as the luminometer. Culture aliquots were removed at regular intervals and assayed with the aldehyde n-decanal in an inexpensive analog luminometer. This luminometer reduces the density of the DNA contents. After this the density of the DNA in the liquid is very much reduced. The content is kept like this for few hours. As hours increase the effect of luminometer becomes higher. But the bacteria should not be in an extreme luminescence state. It will destroy the DNA. So it can accept only to a limit of luminometer. Then the luminometer is removed from the DNA.



**Fig: 8 The gold rods that is inserted in the DNA content. Those rods is of only few nanometer in length**

In the presence of ultraviolet light the liquid medium containing DNA is encoded with digital data. The presence of ultraviolet light is to increase the effect of mutations. By the effect of ultraviolet light the effect of mutation becomes more spontaneous.

Now two electrodes are inserted into this liquid medium which contains the DNA. These electrodes are made of gold. They are of one nanometer in length. Four such electrodes are used in this. The digital data is converted into electrical impulses. These impulses are sent through these electrodes and this will encode the data onto the DNA. After encoding the data on the synthetic DNA, the DNA is inserted into living bacteria as a living medium. This data will live along with the bacteria for generations.

By this method data can be stored in the DNA and data will be transmitted over many generations. These data can last even up to 3000 years. Hence, even after 100 human generations we can get the data. Bacteria may be an inexpensive and stable long-term means of data storage. We can store high content of data on a small piece which can just measure a few nanometers. Genetic coding is so massive that information can be stashed away somewhere in the gene without affecting an organism's overall appearance and other traits.

#### 5. RETRIEVING DATA

Data stored on the DNA can be retrieved by decoding it.[2] The reverse process in the encoding is just done. The DNA is extracted from the bacteria at first. The luminometer is again added and the concentration of DNA is fairly reduced. Electrodes are placed into the liquid medium that contains DNA molecules and ultraviolet light and ionizing radiations are given to the content at 30oC. The encoded data in the DNA get transformed into the electrical impulses and the data are retrieved through the electrodes.

There will not be any loss of information due to this process. The original information that was encoded is retrieved from the content. The amount of information that can be stored depends up on the amount of bacteria in the content. If a milliliter of liquid can contain up to billion bacteria, the potential capacity of such a memory system is enormous. Higher the bacteria content, higher the storage capacity.



## 6. COMPARISON BETWEEN ELECTRONIC AND DNA STORAGE

The main difference is the data stored in DNA can never be lost, but in other storage devices there is always a chance for Data loss.[1][11][9]

Parameter	Digital Storage Device	DNA Storage
Basic Element	Transistor, Magnetic or Optical Domain	Nucleotide
Basic Element Size	1,000 / 10,000 nm <sup>2</sup>	1 nm <sup>3</sup>
Addressing Scheme and Organization	Discrete, Hierarchical	Sequential, Parallel Copying.
Redundancy	Multi Layer	Data Replication
Capacity	Unlimited	0.1 GB / genome
Linear Density	10 kilo bit / mm <sup>2</sup>	1 Mbit / mm
Areal Density	1 GBit / mm <sup>2</sup>	-
Access Time	0.01 milli sec	-
Internal Data Transfer	10 Mbit / sec	100 Kbits / sec
Power Consumption	10 <sup>1</sup> ÷ 10 <sup>2</sup> mW	10 <sup>-8</sup> ÷ 10 <sup>-9</sup>
Error Rate	10 <sup>-4</sup>	10 <sup>-9</sup>

**Table 1: Comparison between DNA storage and Electronic Storage**

## 7. APPLICATIONS

- This biomaterial can be made in small memory sticks that can store data. By this the today's optical memory storage can be replaced by Biostorage Technologies. The device made of this is even smaller than today's USB flash memory "thumbdrives"

- In the optical memory system the range is band limited by the wavelength of light. But here we don't have any band limitations. So we can store enormous amount of data without any band limitations. The modern CD's (Compact Discs) will be replaced by synthetic DNA attached bacteria in the market. That will be more compact and flexible form of memory.

- Each artificial DNA strand can hold up to 100 bits of data and are preserved by making multiple copies of the DNA and inserting the original as well as the identical copies into the bacterial genome sequence. So it is very easy to make multiple copies of the DNA. So it is possible of taking copies of the data in a more efficient manner by the method of cloning. This is more durable than the other data storage mechanisms.

- Boffis an industry that makes research on this says that data can be stored for millions of years. [4] DNA was perhaps the oldest data storage medium and that the work could lead to text, images, music and other digital data being stored inside the genomes of living organisms

- We can spot the sense of living organisms in places. This will be helpful to identify bodies struck in places due to disasters

- Other applications are to create living data storage for nano-computers, or to tag an organism for life using a unique identifier strand inserted into the DNA. Computer memory like the magnetic disc, semiconductors, superconductors, etc may fail due to lots of problems. It can be affected by external radiations. But the data stored here is more reliable.

## 8. RELIABLE BACTERIAL DATA STORAGE

This simple, flexible and robust method offers a practical solution to data storage and retrieval challenges in combination with other previously published techniques.[12] One early use for the technique would be to create special markers to identify legitimate versions of pharmaceuticals. However, the bacillus itself creates new copies of the data every time it reproduces itself, thus making it an ideal archival storage system. Each artificial DNA strand can hold up to 100 bits of data and preserved by making multiple copies of the DNA and inserting the original as well as the identical copies into the bacterial genome sequence. These copies work as back-ups of the data to counteract natural degradation.

Current data storage only lasts around 100 years, but this technology could allow the safe storage of data for millions of years. Information storage using DNA is robust for more than one hundred million years. According to the researchers, bacteria have particularly compact DNA which is passed down from generation to generation. Although mutations do occur as the data is passed from generation to generation, the rate should be slow enough to maintain the data integrity. The artificial DNA that carries the data to be preserved makes multiple copies of the DNA and inserts the original as well as identical copies into the bacterial genome sequence. The multiple copies work as backup files to counteract natural degradation of the preserved data, according to the newswire. The information stored in that DNA can also be passed on for long-term preservation of large data files, the scientists said.

The four characters that represent the genetic coding in DNA work much like digital data. Character combinations can stand for specific letters and symbols — so codes in genomes can be translated, or read, to produce music, text, video and other content. Genetic coding is so massive that information can be stashed away somewhere in the gene without affecting an organism's overall appearance and other traits. But mutation could distort stored data. So it is stored in four DNA strands in a bacteria.

## 9. CONCLUSION

From the present study the following conclusions can be drawn. Bacteria are the future data storing device. Data can be stored in bacterial cells and can be retrieved after 1000 years. Bacteria are cheap and store enormous data. All future technologies may be based on the data storing in biological material including bacterial cells.

## 10. REFERENCES

- [1] Digital and Biological Storage Systems - a Quantitative Comparison, M. Wasizu, O. Kurosawa, I. Arai, S. Suzuki, and N. Shimamoto, IEEE 2005.
- [2] Quantitative Analysis of DNA Orientation in Stationary AC Electric Fields Using Fluorescence Anisotropy Seiichi Suzuki, Member, IEEE, Takeshi Yamanashi, Shinichi Tazawa, Osamu Kurosawa, and Masao Washizu, Senior Member, IEEE 2008
- [3] National Center for Biotechnology Information.
- [4] Boffins store data in bacterial DNA, by Ian Williams
- [5] TRN, the Latest Technology Research News.
- [6] NewScientist.com, Data stored in multiplying bacteria
- [7] PhysOrg.com, DNA For Information Processing and Data Storage
- [8] FORESIGHT foresight.org nanotech institute, Data storage using DNA.
- [9] LiveScience.com New Technique Stores Data in Bacteria by Bill Christensen
- [10] Researchers store data in bacteria DNA. Information lasts as long as a species stays alive — possibly a million years, msnbc.msn.com
- [11] R. Staden, "A new computer method for the storage and manipulation of DNA gel reading data", MRC Laboratory of Molecular Biology Hills Road, Cambridge CB2 2QH, UK, Oxford Journals, June 16, 1980.
- [12] Be'atrice Godard, Jo'rg Schmidtke, Jean-Jacques Cassiman and Se'gole'ne Ayme', "Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective", INSERM SC11, Hopital Broussais, Paris, France; Institute of Human Genetics, Hannover Medical School, Hannover, Germany; and Center for Human Genetics, Leuven, Belgium. 2003
- [13] Roger Staden, "Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing", Oxford Journals, 1982
- [14] C T Chung, S L Niemela, and R H Miller, "One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution", Proceedings of the National Academy of Science of the United States of America, April 1989