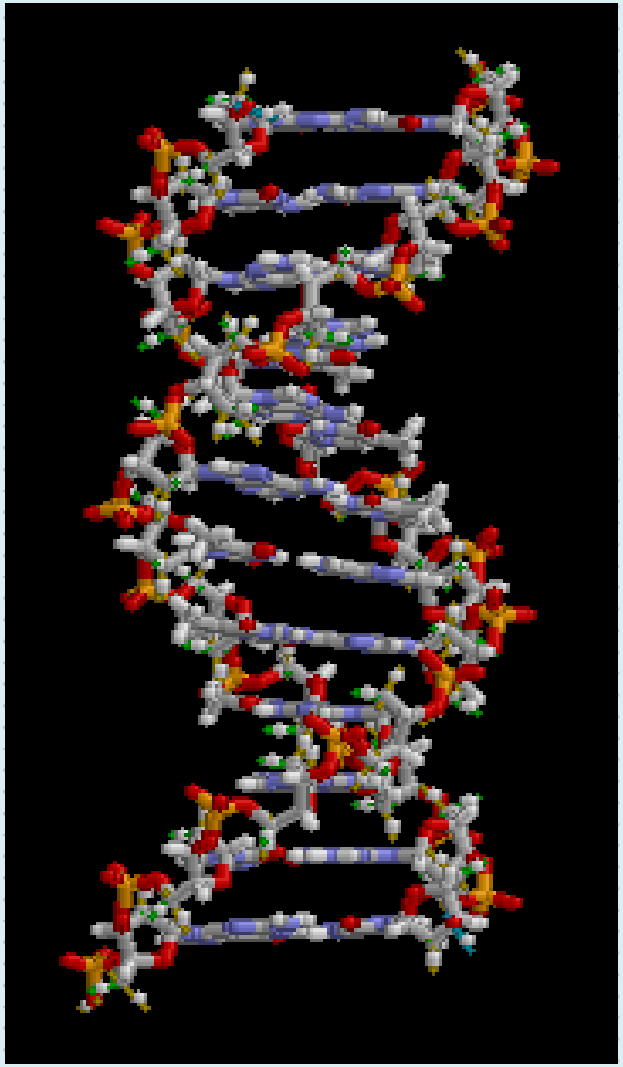
DNA DIGITAL DATA STORAGE

By : Deepak Raya
160115735091

The DNA

- The DNA molecule is a double-helix staircase of billions of molecular blocks, called base pairs.
- Whose arrangement determines much of what makes each of us unique.



What is DNA data storage?

- DNA is nature's hard drive.
- The process of encoding and decoding binary data to and from synthesized DNA strands.
- The basic idea is to convert the data into DNA alphabets (A,T,G,C).

Origin of this idea

- DNA data storage became a popular topic in the 21st century.
- Its origins date back to 1964-65
- When Mikhail Neiman wrote about the possibilities of storing information in DNA.
- Ewan Birney and Nick Goldman(2012).

Motivation for this idea

- Conventional mass-storage systems were doing the job cheaply and reliably.
- The situation has changed drastically over the last 15 years.
- We face an unprecedented data deluge.
- Started with problem of compact and efficient storage system.

Why in DNA ?

- Data density and size advantage:

1. Lot of data can be stored into tiny bit of spaces .
2. 1 g of DNA → 216 petabytes of data
3. Which is not the case with conventional mass storage systems.

- Shelf life advantage:

1. Normal conditions half life is around 500 years
2. In dark and cold conditions its arounds thousands of years

- The power advantage:

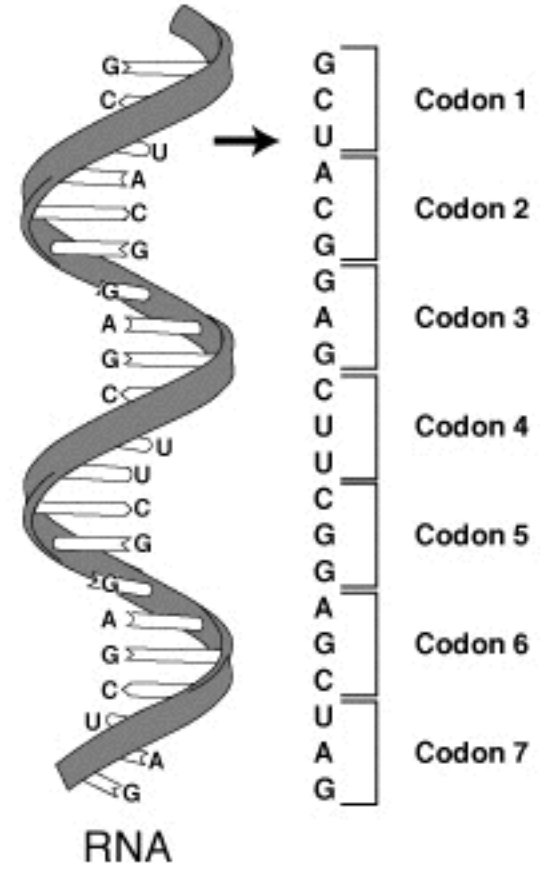
1. Unlike the conventional mass storage systems, it doesn't require the electric power to maintain the data .

- Thus, a compact and efficient storage system.

Big picture of DNA storage system

Cell instructions inside DNA :

- DNA made of four organic bases .
- The specific sequence of these bases into groups of three called codons.
- Codons → instruction to make Proteins.

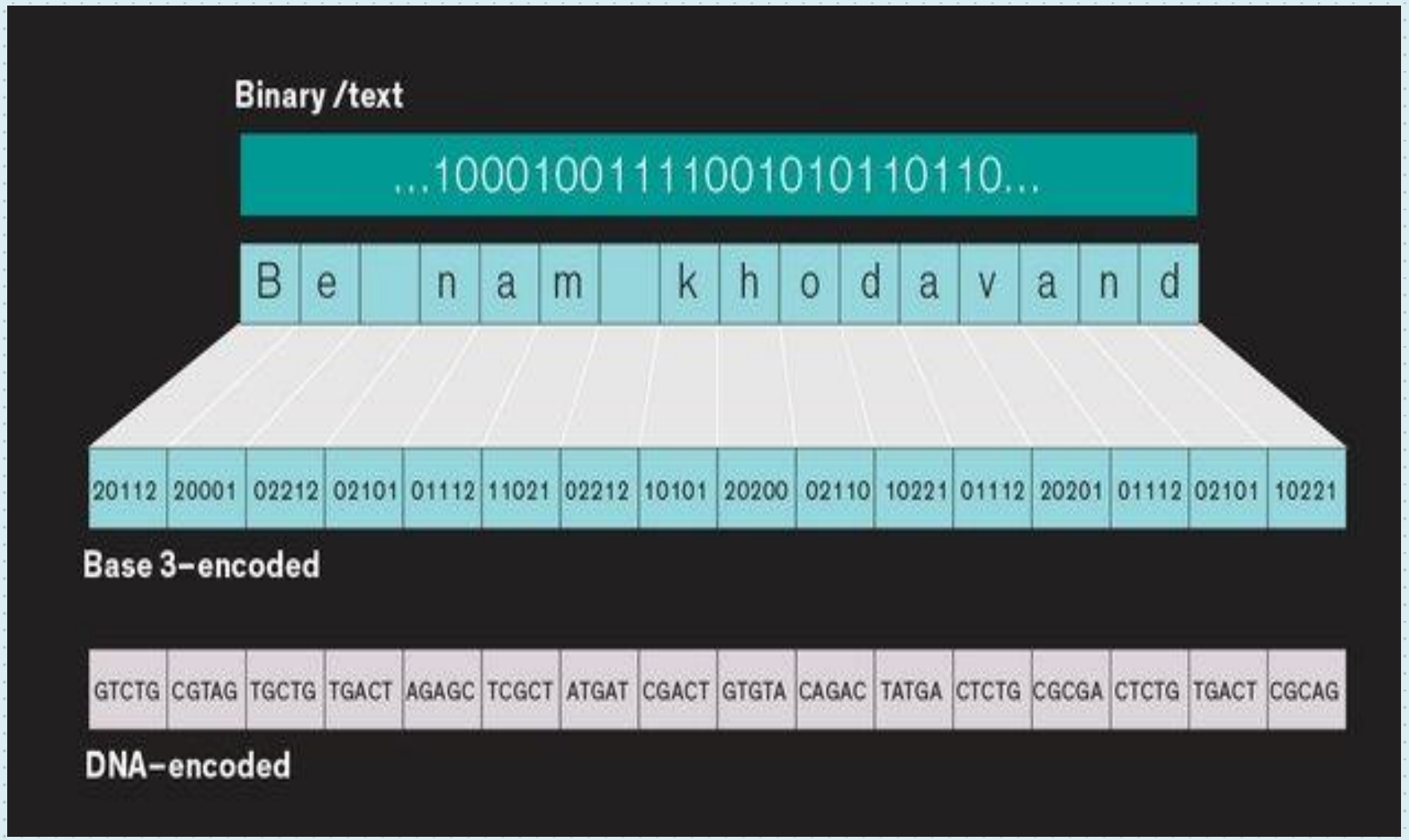


Now, for digital data.

- The codon kind of codes can be used to store other information too.
- Binary encoding → base 3 encoding → DNA encoding.
- Splice the encoded message into DNA strand.

Retrieving information:

- The encoded DNA strand is surrounded with genetic markers.
- The above stand is sequenced using PCR and the message is decoded.
- And if we have few redundant copies of information ,we can correct mistakes automatically.


Binary /text
...10001001111001010110110...
B e n a m k h o d a v a n d
20112 20001 02212 02101 01112 11021 02212 10101 20200 02110 10221 01112 20201 01112 02101 10221
Base 3-encoded
GTCTG CGTAG TGCTG TGACT AGAGC TCGCT ATGAT CGACT GTGTA CAGAC TATGA CTCTG CGCGA CTCTG TGACT CGCAG
DNA-encoded

Coding theory and DNA storage

- DNA-based storage systems are new and uncharted territory for coding theorists.
- As the amount of data to store increase we need to augment with error control coding.
- Types of encoding:

1. Base 3 encoding.
2. 2 bits per base encoding.

- **Ensuring redundancy in DNA:**

1. Encoding the data in shorter strands.
2. Construct the first part of the next strand using the same data found at the end of the previous strand.
3. multiple copies of the data for comparison.

- **Errors possible:**

1. Substitution
2. Swapping

DNA sequencing

- Reading out the data stored.
- The cost for sequencing reduced a lot.

- Plenty of devices.
- Various methods:

1. Shotgun sequencer
2. Nano pore-sequencer

- Different sequencers-Different codes.

Downsides of DNA data storage

- Cost: It involves high cost to synthesize and sequence DNA.
- Time: There is a delay in process of synthesis and sequencing DNA.
- Complexity: DNA isn’t so simple. It’s inherently unordered and lot of things to consider.

Few excerpts of storing information into DNA

- 2012-scientist encoded 739 KB of computer files-154 Shakespeare sonnets and Martin Luther King Jr. “I have a dream” speech.
- 2016-Microsoft research and University of Washington- stored 200MB of data

Conclusion

- DNA based storage system can act as compact and efficient storage systems.
- Although the cost of DNA synthesizing and sequencing is high, it has reduced to one three-millionth what it was 10 years ago.
- There are already few excerpts of data storage in DNA with zero error in retrieval.
- There is a lot of research going behind DNA storage systems, probably in near future we might store the entire internet data (1.1 ZB) into few grams of DNA.

References

- Olgica Milenkovic. (2018)- “Exabytes in a Test Tube: The Case for DNA Data Storage”- IEEE spectrum.
- Goldman, N. B.; Birney, E. (2013). "Towards practical, high-capacity, low-maintenance information storage in synthesized DNA". *Nature*. **494** (7435): 77–80.

Thank you !