

DNA as a repository of information

MT-202

Zakirov



- Computer data can be transferred to the basis of life itself-DNA. DNA molecules can potentially store the entire volume of the world's digital information — already 1.1 zettabytes (10^{21}) of data - in about 9 liters of solution for millennia.



18 litre
bottle



One grain of sand = one exabyte

1 Exabyte = 1073741824 Gigabytes

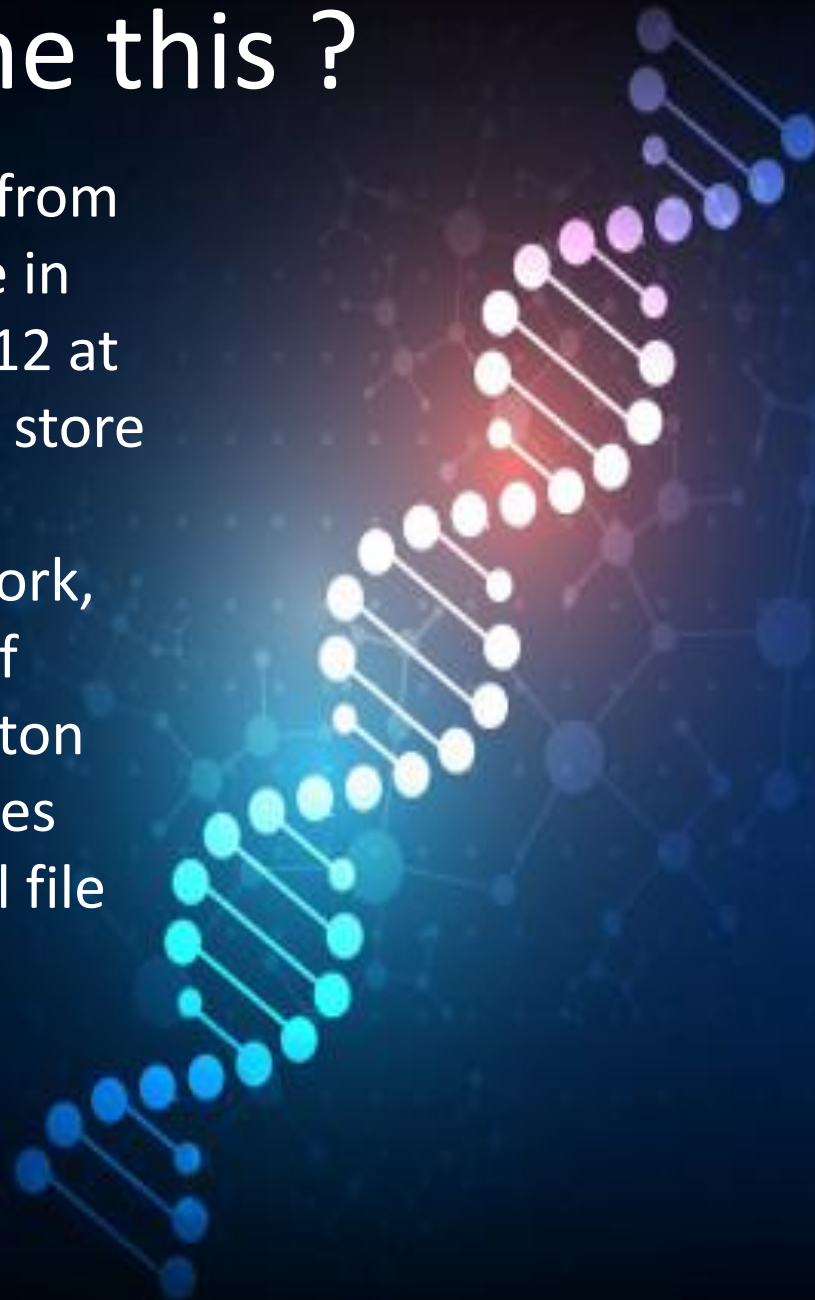


- The information encoded in DNA in exabytes of data can theoretically be stored in the volume that takes up one grain of sand.



Have people ever done this ?

- Experiments conducted by scientists from the European bioinformatics Institute in Hinxton (England) in 2013 and in 2012 at Harvard, showed that it is possible to store data files in DNA, and then read the information digitally. Based on this work, research teams from the University of Illinois and the University of Washington were able to save four small image files and then restore them using a special file identifier.



Anyone else?

- In 2016, researchers from Microsoft and the University of Washington managed to do what no one else had been able to do before — record 200 MB of data in the form of a sequence of nucleotides that make up artificially created DNA. It is already quite close to the 750 MB contained in human DNA.



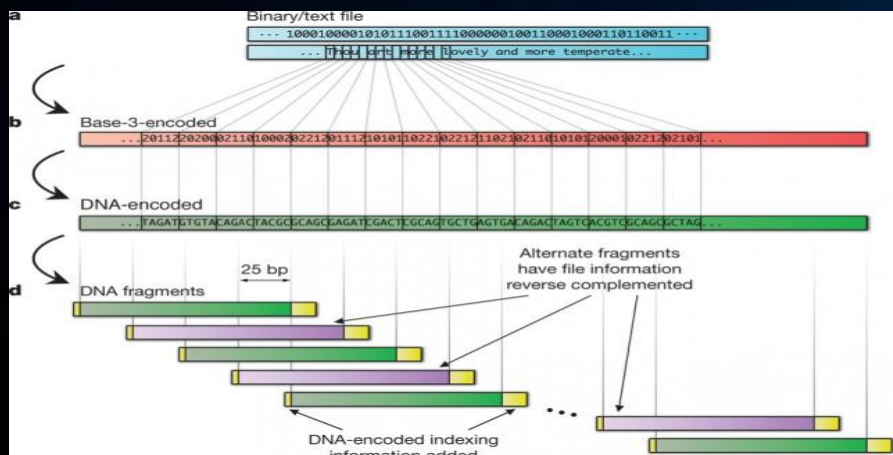
Microsoft

DNA?

giving
ving
I
ead
ary

A stylized DNA double helix structure is depicted, composed of glowing blue and white spheres connected by lines, set against a dark blue background with a subtle molecular pattern.

-
- The diagram illustrates the process of encoding a binary/text file into DNA fragments for storage and retrieval, showing the flow from a binary/text file to DNA fragments and back to the original data.
- a. Binary/text file:** The input is a binary/text file containing a sequence of bits (e.g., 10001000010101100111000000100110001000110110011...) and a corresponding text string (e.g., "... THE HILL IS LOVELY and more temperate...").
 - b. Base-3-encoded:** The binary data is converted into a Base-3-encoded sequence (e.g., ...201122020002211010000222120111210101102210222121102102110101020001022102101...).
 - c. DNA-encoded:** The Base-3-encoded sequence is converted into a DNA sequence (e.g., ...TAGATGTGTACAGACTACGGCCAGCCAGATCGACTCGCACTGCTGAGTGACAGACTAGTCACGTCGCCAGCCCTAG...).
 - d. DNA fragments:** The DNA sequence is fragmented into 25 bp (base pairs) segments. These fragments are stored in a database, and the original data is reconstructed by reverse complementing the fragments and indexing them.



Why is it not actively used today?

- Equipment for working with DNA is prohibitively expensive (the cost of encoding information in DNA is estimated at about \$ 12,400 per megabyte, the cost of reading - \$ 220 for 1 megabyte), but the cost of sequencing or "reading" the genetic code falls faster than the cost of computer memory, and technologies for creating synthetic DNA continue to develop. But another important problem has not been solved — the speed of writing and reading information takes several hours.



Result

- DNA is a great medium for long- term storage. If in the first experiments it was necessary to maintain cold and dryness, then in subsequent experiments the information could be stored at room temperature. And if you add DNA to a quartz ball and store it at -18°C , the information will be stored for millions of years.



THANK YOU

IT'S ALL

