BEYOND BYTES: THE ERA OF DNA-ENCODED INFORMATION

Kachave Mukund Ramesh[1], Komal[2]

[1] PhD scholar, Department of Animal Genetics and Breeding, ICAR-NDRI, Karnal, Haryana-132001, India

[2] Assistant Professor, Department of Animal Nutrition, Mahatma Gandhi Veterinary College, Bharatpur, Rajasthan-321001, India

Corresponding author email: mukundrkachvae@gmail.com

DOI: https://doi.org/10.5281/zenodo.15221163

Abstract

DNA data storage is an innovative technology that encodes digital information into DNA sequences, offering unmatched density, durability, and sustainability. The process involves converting binary data into nucleotide sequences, synthesizing DNA, storing it under stable conditions, and retrieving information through sequencing and decoding. DNA storage provides long-term stability, minimal space requirements, and energy efficiency, making it ideal for archival purposes. However, challenges such as high costs, slow read/write speeds, and sequencing errors limit its widespread adoption. Potential applications include historical data preservation, biomedical research, and space exploration. Ongoing advancements in automation and cost reduction strategies aim to enhance feasibility. With continued research, DNA storage could revolutionize digital data preservation for future generations.

Keywords: DNA sequences, DNA data storage, biomedical research, space exploration

Introduction

With the exponential growth of digital data, conventional storage technologies such as hard drives, magnetic tapes, and solid-state drives are reaching their limits in terms of capacity, longevity, and sustainability. DNA, the molecule that stores genetic information in all living organisms, is emerging as a promising alternative for data storage. Its remarkable density, stability, and longevity make it a viable candidate for archiving vast amounts of information in a compact and durable form. This article explores the mechanism behind DNA data storage, its advantages and disadvantages, and its potential impact on future information storage.

How DNA Data Storage Works

DNA data storage is based on the principle of encoding digital information (binary code) into the four nucleotide bases of DNA: adenine (A), cytosine (C), guanine (G), and thymine (T). The process involves several key steps:
DNA data storage is an emerging technology that encodes digital information into DNA sequences. The process involves several key steps:

1. Binarization

The first step in storing data in DNA is converting the original data (text, images, videos, etc.) into a binary format (0s and 1s). Since computers process information in binary code, this transformation ensures compatibility with digital data storage systems.

2. Encoding

Once the data is in binary format, it must be converted into a DNA sequence. This is achieved by mapping binary values to the four nucleotides of DNA:

- **A (Adenine)**
- **T (Thymine)**
- **C (Cytosine)**
- **G (Guanine)**

Different encoding schemes are used to ensure error correction and stability. One common approach is assigning two-bit binary combinations to nucleotides, such as:

- 00 → A
- 01 → C
- 10 → G
- 11 → T

Advanced encoding techniques, such as error-correcting codes (Reed-Solomon codes or Hamming codes), are used to minimize mutations and sequencing errors.

3. DNA Synthesis

After encoding, the digital data is synthesized into actual DNA molecules. This is done using chemical or enzymatic synthesis, where short DNA oligonucleotides are created based on the encoded sequence. DNA synthesis technologies, such as phosphoramidite chemistry, enable the creation of custom DNA strands with high accuracy.

4. Storage

Once synthesized, the DNA strands are stored in a stable environment. DNA is highly durable and can be preserved for thousands of years under optimal conditions, such as dry, cold environments. Unlike traditional storage media, DNA does not degrade quickly and has an extremely high data density.

5. Sequencing (Reading Data)

To retrieve stored information, the DNA is sequenced using next-generation sequencing (NGS) or nanopore sequencing methods. These techniques read the DNA strands and convert them back into digital nucleotide sequences.

6. Decoding and Reconstruction

The final step involves translating the sequenced nucleotide data back into binary format using the predefined encoding scheme. Error correction algorithms help recover lost or damaged data. Once the binary data is restored, it is converted into its original format (text, image, video, etc.).

Advantages of DNA Data Storage

1. Unprecedented Data Density

DNA offers an unmatched storage density. One gram of DNA can theoretically store up to 215 petabytes (215 million gigabytes) of data, making it far superior to traditional storage devices.

2. Longevity and Durability

Unlike conventional digital storage media, which degrade over time, DNA can remain stable for thousands of years when stored properly. Fossilized DNA has been retrieved from ancient organisms, demonstrating its resilience.

3. Minimal Space Requirement

DNA's high density means that entire data centre's worth of information can be stored in a few grams of DNA, reducing the physical space required for data storage.

4. Sustainability and Energy Efficiency

DNA storage does not require constant power consumption for maintenance, unlike data centre's that demand significant energy for cooling and operation. This makes it an environmentally friendly alternative.

5. Non-Obsolescence

Unlike magnetic tapes, CDs, and hard drives, which become obsolete with evolving technology, DNA remains a universal biological medium that will always be relevant.

Challenges and Disadvantages

1. High Cost

The synthesis and sequencing of DNA are currently expensive. Although costs have significantly decreased in recent years, they remain a barrier to large-scale adoption.

2. Slow Read and Write Speeds

- Writing (synthesizing) and reading (sequencing) DNA is significantly slower than electronic storage devices. This makes it impractical for real-time data access but ideal for long-term archival storage.

5. Non-Obsolescence

Unlike magnetic tapes, CDs, and hard drives, which become obsolete with evolving technology, DNA remains a universal biological medium that will always be relevant.

Challenges and Disadvantages

1. High Cost

- The synthesis and sequencing of DNA are currently expensive. Although costs have significantly decreased in recent years, they remain a barrier to large-scale adoption.

2. Slow Read and Write Speeds

- Writing (synthesizing) and reading (sequencing) DNA is significantly slower than electronic storage devices. This makes it impractical for real-time data access but ideal for long-term archival storage.

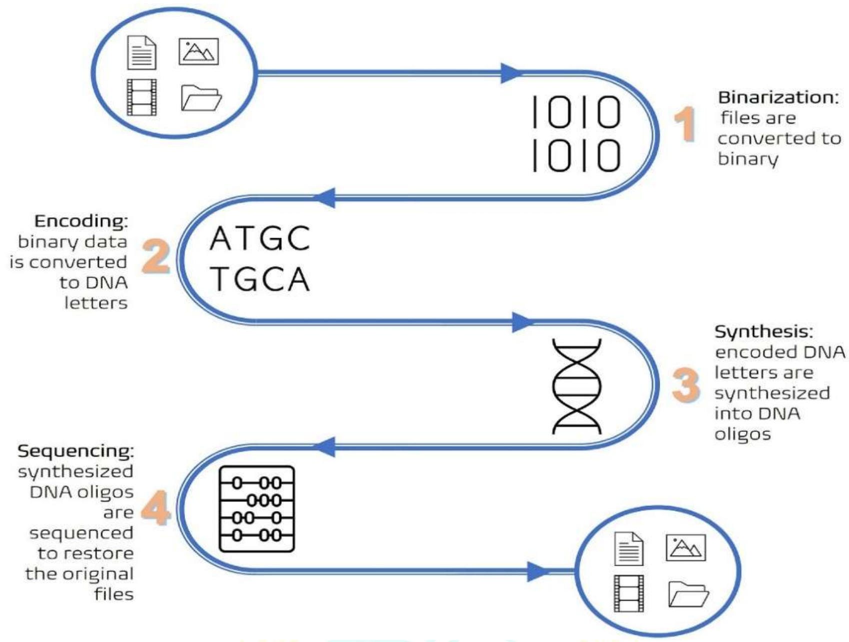

Fig: Steps of Data Storage in DNA

Conclusion

DNA-based data storage presents an exciting frontier in digital information preservation. Its exceptional density, longevity, and environmental sustainability make it a strong contender for the future of data storage. While challenges such as cost and read/write speeds remain, ongoing advancements in DNA synthesis and sequencing hold promise for overcoming these barriers. As research progresses, DNA storage may revolutionize how humanity stores and preserves information for generations to come.

References

Akash, M., Bencurova, E., & Dandekar, T. (2024). How to make DNA data storage more applicable. *Trends in Biotechnology, 42*(1), 17–30. https://doi.org/10.1016/j.tibtech.2023.07.006

Bar-Lev, D., Orr, I., Sabary, O. *et al.* Scalable and robust DNA-based storage via coding theory and deep learning. *Nat Mach Intell* (2025). https://doi.org/10.1038/s42256-025-01003-z

Ceze, L., Nivala, J. & Strauss, K. Molecular digital data storage using DNA. *Nat Rev Genet* **20**, 456–466 (2019). https://doi.org/10.1038/s41576-019-0125-3

Sensintaffar, A., Wei, Y., Ou, L., Du, D., & Li, B. (2025).

Advancing archival data storage: The promises and challenges of DNA storage systems. *ACM Transactions on Storage, Just Accepted (March 2025).* https://doi.org/10.1145/3723166

Yang, S., Bögels, B.W.A., Wang, F. *et al.* DNA as a universal chemical substrate for computing and data storage. *Nat Rev Chem* **8**, 179–194 (2024). https://doi.org/10.1038/s41570-024-00576-4

Cite this article at:

Kachave Mukund Ramesh, Komal. (2025). BEYOND BYTES: THE ERA OF DNA-ENCODED INFORMATION. Vet Farm Frontier, 02(03), 7–9. https://doi.org/10.5281/zenodo.15221163