



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 6, June 2024

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

Impact Factor: 8.423

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)

# Advancements in DNA Data Storage: Scaling, Retrieval, and Selective Reading

Saksham Batta, Ankita Badhe

Department of MCA, Vivekanand Education Society's Institute of Technology (VESIT), Chembur, Mumbai, India

Department of MCA, Vivekanand Education Society's Institute of Technology (VESIT), Chembur, Mumbai, India

**ABSTRACT:** DNA data storage represents a groundbreaking approach to preserving vast amounts of information in a compact, durable form. This paper synthesizes findings from two major studies, one by Lee Organick et al. and another by Elshaika, to present a unified perspective on advancements in DNA data storage. Key topics include scaling up storage capabilities, improving random access retrieval, implementing robust error correction mechanisms, and developing selective data reading techniques. This integrated approach highlights the progress and potential of DNA as a long-term data storage medium.

## I. INTRODUCTION

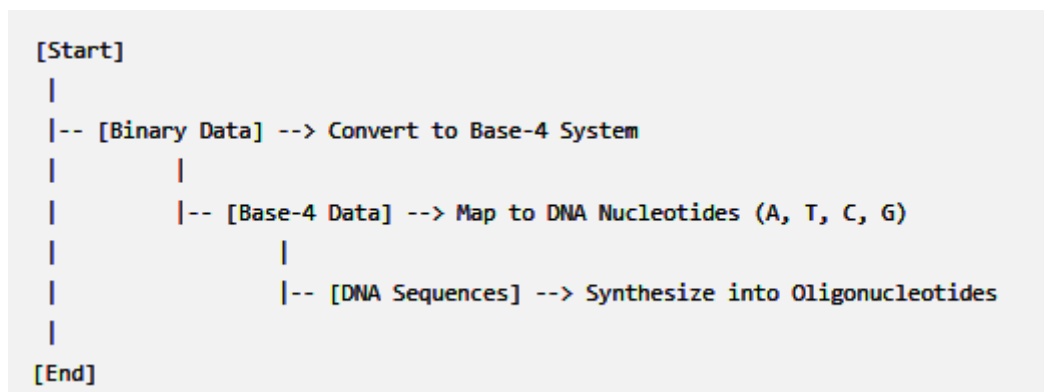
The digital era has seen an unprecedented explosion in data generation, necessitating novel storage solutions that surpass the capabilities of traditional media. DNA, with its exceptional density and stability, offers a promising alternative. This paper integrates the findings of two significant studies: one focused on scaling and random access in DNA storage, and the other on selective reading of DNA-stored data. Together, these studies provide a comprehensive overview of the current state and future directions of DNA data storage technology.

### Scaling Up DNA Data Storage

- **Data Encoding and Storage**

The study by Lee Organick et al. has made substantial strides in encoding and storing large volumes of data in synthetic DNA. They encoded 200MB of data into 13 million DNA oligonucleotides, achieving error-free data recovery. This milestone demonstrates the feasibility of scaling DNA data storage for practical applications. The data encoding process involved converting binary data into a base-4 system, compatible with the nucleotide sequences of DNA (A, T, C, G), which were then synthesized into oligonucleotides.

Diagram 1: Data Encoding Process

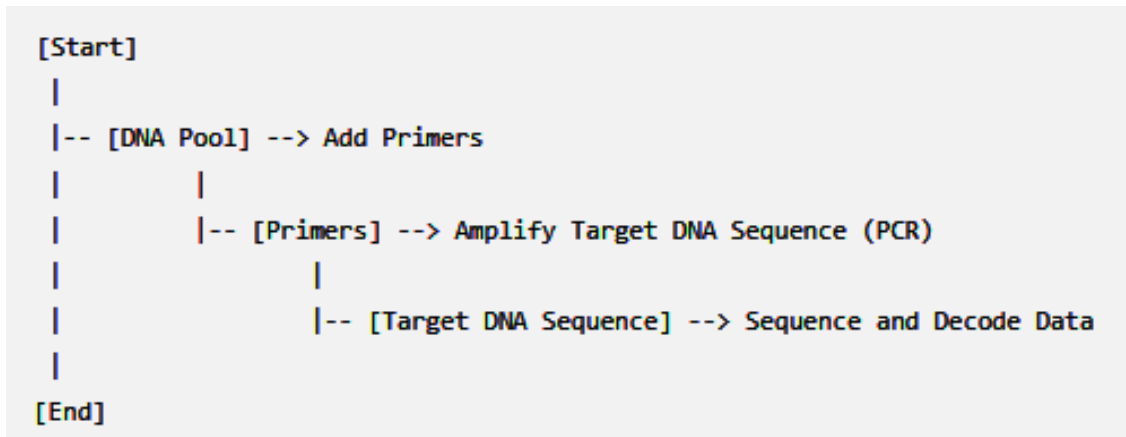


- **Random Access Retrieval**

Efficient random-access retrieval is essential for the practical use of DNA as a storage medium. Organick et al. developed a polymerase chain reaction (PCR)-based methodology, utilizing a large library of primers to target specific

data sequences within the DNA pool. This method enables the retrieval of specific files from a large collection of stored data, illustrating the potential for practical, large-scale DNA data storage solutions.

Diagram 2: Random Access Retrieval Process



• **Error Correction and Physical Redundancy**

DNA synthesis and sequencing are inherently error-prone processes, with common errors including insertions, deletions, and substitutions. To mitigate these issues, Organick et al. introduced a novel coding scheme that reduces physical redundancy while maintaining logical density. This approach ensures robust data recovery, even in the presence of high error rates, making DNA a more reliable storage medium.

Chart 1: Error Rates in DNA Synthesis and Sequencing

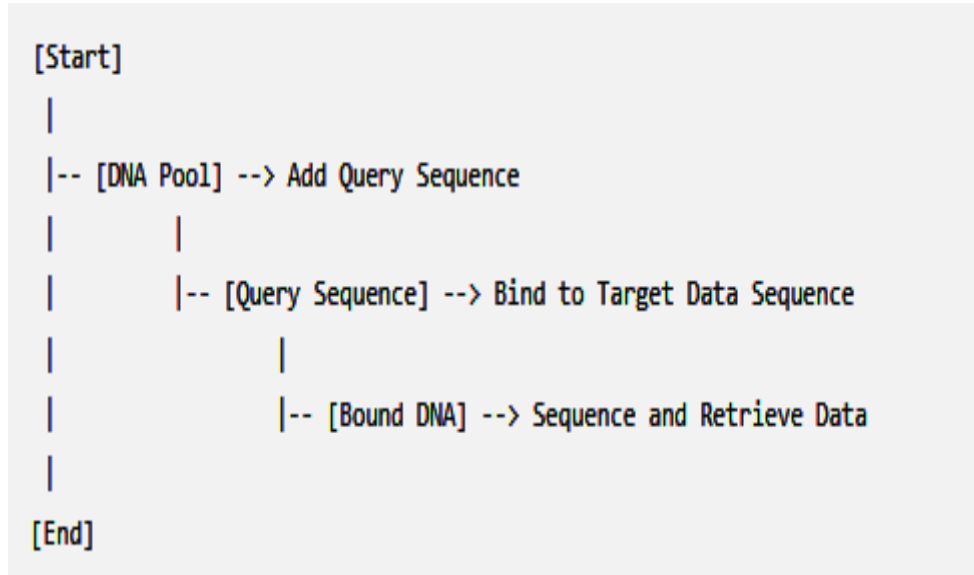
Error Type	Rate (%)
Insertions	1.5
Deletions	0.8
Substitutions	0.2
<b>Total Error Rate</b>	<b>2.5</b>

Selective Data Reading

• **Content-Based Queries**

Elshaika’s research focuses on the selective reading of DNA-stored data, an essential feature for efficient data management. This study explores methods for implementing key-value and content-based queries directly on DNA, without the need for additional storage devices. The ability to perform selective reading directly on DNA enhances the practicality and cost-effectiveness of using DNA as a storage medium.

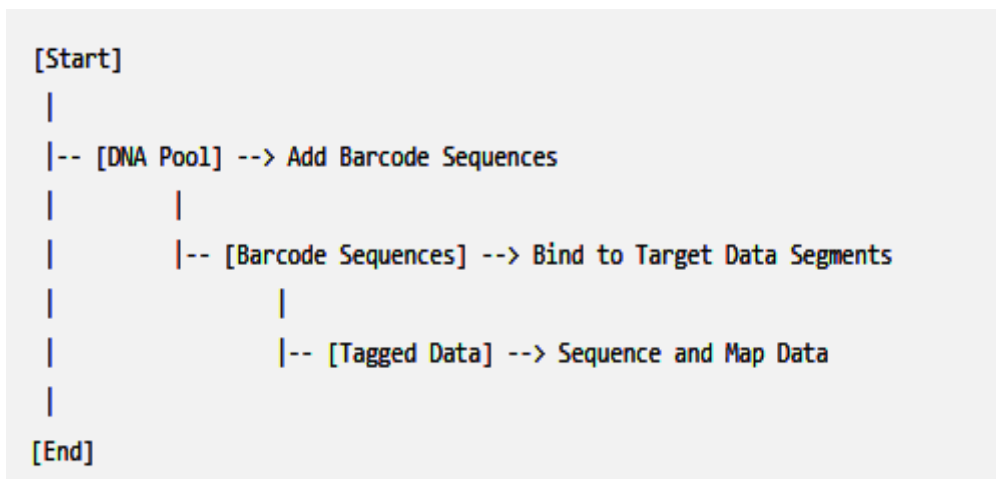
Diagram 3: Content-Based Query Process



- **Barcode-Based Random Access**

A notable innovation from Elshaika's study is the use of barcodes for managing and querying data within a DNA pool. Barcodes enable key-value queries and potentially support content-based queries through in-silico mapping for unique attributes. This system facilitates the efficient organization and retrieval of data, further advancing the practical application of DNA data storage.

Diagram 4: Barcode-Based Random-Access Process



## II. DISCUSSION

The integration of scalable storage techniques and selective data retrieval methods marks a significant advancement in DNA data storage. The combination of large-scale data encoding, robust error correction, and efficient random-access retrieval, as demonstrated by Organick et al., sets a strong foundation for practical DNA data storage solutions. Additionally, the selective reading techniques developed by Elshaika enhance the usability of DNA as a storage medium, making it possible to perform complex queries directly on the stored data.



#### IV. CONCLUSION

The studies by Organick et al. and Elshaika collectively enhance our understanding of the potential and challenges of DNA data storage. By combining scalable encoding and storage techniques with innovative retrieval and error correction methods, these advancements pave the way for the practical application of DNA as a long-term data storage medium. Future research should continue to explore ways to improve the efficiency and reliability of DNA data storage, ensuring its viability as a solution for the growing data storage demands.

#### REFERENCES

1. Lee Organick, Siena Dumas Ang, Yuan-Jyue Chen, et al. "Scaling up DNA data storage and random-access retrieval." University of Washington, Microsoft Research.
2. Elshaika. "Selective Reading of DNA-Stored Data." University of Marburg, Scientific Reports.



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  [ijirset@gmail.com](mailto:ijirset@gmail.com)



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details