

DNA Data Storage: Multidisciplinary Approaches to Long-Term Digital Preservation Towards Advancements, Challenges, Future Prospects

Zarif Bin Akhtar^{1*}, Ahmed Tajbiul Rawol²

¹Department of Computing, Institute of Electrical and Electronics Engineers, United States.

²Department of Computer Science, American International University-Bangladesh, Bangladesh.

*Corresponding author: Zarif Bin Akhtar.

Abstract

This research exploration provides a very comprehensive analysis of DNA-based data storage systems, focusing on their capacity to address the increasing demand for long-term, sustainable data archival. Through an investigative iterative exploration of DNA computing principles, accelerated computing inclusive of artificial intelligence (AI) along with various types of the many recent breakthroughs within synthetic biology in terms of both biomedical engineering, systems biology and unconventional computing methodologies, we evaluate towards the available, associated capabilities and limitations of DNA storage. Special emphasis is also placed on the impacts of interdisciplinary collaborations within biomedical engineering, cognitive computing, computational mechanics and systems biology. Our findings demonstrate the viability of DNA data storage as an efficient and scalable solution for the current digital age, contributing to the advancement of data management technologies.

Keywords: artificial intelligence; biomedical engineering; biotechnology; data archival; deep learning; DNA data storage; machine learning; systems biology

Introduction

The digital age is drowning in data. Traditional electronic storage is reaching its limits, demanding innovative, sustainable solutions. Enter DNA. This research explores DNA's potential as a revolutionary data storage medium, leveraging its inherent information encoding capabilities for long-term archival [1-3]. The exponential growth of data from cloud computing to IoT necessitates storage with unparalleled durability, compactness, and longevity [4-6]. By merging synthetic biology, unconventional computing, and DNA-based systems, we are witnessing a paradigm shift. This analysis delves into the advancements, challenges, and future directions of DNA data storage, aiming to illuminate its transformative impact. We examine the current state-of-the-art techniques and theoretical frameworks, providing a roadmap for its practical integration. DNA offers superior efficiency and longevity, with better compression, higher density, and lower energy costs. However, crucial considerations like metadata integration, bio cybersecurity, and standardization must be addressed. Robust security protocols, standardized coding/decoding, and reliable handling procedures are essential. Integrating identifying markers within the DNA structure ensures future accessibility [7-9]. While challenges remain regarding

scalability and specialized tools, interdisciplinary collaboration is key to unlocking DNA's full potential for high-density, long-term, and energy-efficient data storage.

Methods and Experimental Analysis

This investigative analysis employed a multi-faceted methodology to comprehensively evaluate DNA data storage systems. The mixed layered approach encompassed a systematic exploration of existing available knowledge, primary data collection, expert consultations, comparative analysis, and the development of a strategic roadmap.

Background Research Explorations

Comprehensive background research for available knowledge was analyzed and investigations were conducted, encompassing peer-reviewed research articles, technical reports, and scholarly publications related to DNA data storage, synthetic biology, and unconventional computing.

Databases such as PubMed, IEEE Xplore, and Google Scholar were utilized, employing keywords related to DNA data storage, DNA synthesis, encoding/decoding algorithms, and bio cybersecurity. The analysis aimed to establish a thorough understanding of the current state-of-the-art, identify

key research trends, and synthesize a robust knowledge base for subsequent analysis.

Primary Data Collection and Analysis

Empirical data and case studies focusing on the implementation of DNA data storage systems were systematically collected. This involved analyzing experimental results from published studies, focusing on metrics such as data density, retrieval accuracy, synthesis/sequencing costs, and long-term stability. Both qualitative and quantitative methods were applied to analyze the collected data.

Quantitative Analysis: Statistical analysis was performed to identify trends and patterns in data, using metrics like data density, error rates, and storage capacity.

Qualitative Analysis: Case studies were examined to identify challenges and best practices in DNA data storage implementation, focusing on factors like encoding strategies, error correction, and data retrieval methods.

Expert Consultations

In-depth interviews and consultations were conducted with leading experts in molecular biology, computational genomics, and synthetic biology. These consultations provided insights into the feasibility, scalability, and ethical implications of DNA data storage.

Semi-structured interviews were conducted, focusing on key areas such as:

- Current limitations of DNA data storage technologies.
- Potential solutions for scalability and standardization.
- Bio cybersecurity considerations.
- Future directions for research and development.

Comparative Analysis

A comparative analysis was performed to evaluate the efficacy and efficiency of DNA data storage systems against conventional electronic storage methods.

Multiple case studies were examined, comparing performance metrics such as:

- Storage capacity and density.
- Data retrieval speed and accuracy.
- Long-term stability and durability.
- Energy consumption and cost.

This analysis allowed for a thorough assessment of the advantages and limitations of DNA data storage. Environmental variables were considered in the

analysis of data stability, such as temperature, and humidity.

Roadmap Development

Based on the insights gained from the investigative explorations, data analysis, expert consultations, and comparative studies, a comprehensive roadmap was developed. This roadmap outlined key research priorities, technological advancements, and interdisciplinary collaborations necessary to address existing challenges and facilitate the practical integration of DNA data storage.

The roadmap included recommendations for:

- Standardization of encoding and decoding protocols.
- Development of robust bio cybersecurity measures.
- Advancements in DNA synthesis and sequencing technologies.
- Exploration of novel metadata integration strategies.
- Promotion of interdisciplinary collaboration.

Background Research and Investigative Explorations for Available Knowledge

The burgeoning field of DNA digital data storage represents a paradigm shift in information technology, offering the potential for ultra-high-density and long-term data archival. This section provides a detailed overview of the foundational concepts, historical milestones, and recent advancements in this domain.

Foundational Concepts

DNA as a Storage Medium

The core concept involves encoding binary digital data into synthetic DNA strands, leveraging the inherent information storage capacity of DNA's nucleotide sequence (adenine, guanine, cytosine, thymine). While DNA offers exceptional storage density, practical limitations such as cost and speed of synthesis and sequencing remain significant challenges [1-5].

Encoding and Decoding Techniques

Various encoding strategies exist, including:

- Translation of text into DNA codons, analogous to biological protein synthesis [6].
- Conversion of arbitrary binary data into ternary or other base systems for DNA encoding [7,8].
- Error correction mechanisms are essential to maintain data integrity during synthesis, storage, and sequencing, ensuring accurate retrieval.

Historical Milestones

Early Conceptualization

The idea of using DNA for data storage emerged in the mid-20th century, with early theoretical considerations in the 1950s and 1960s.

Proof-of-Concept Demonstrations

Significant milestones include successful encoding of images and entire books into DNA, demonstrating the feasibility of the concept [9-11]. Development of error-correcting codes improved data retention accuracy, addressing a critical challenge.

Recent Advancements

Sophisticated encoding algorithms and automated systems for DNA synthesis and sequencing have improved efficiency and scalability. Integration of DNA storage into living organisms through synthetic biology and CRISPR-Cas9 gene editing offers novel possibilities [12,13].

Emerging Research Areas

In Vivo DNA Data Recording

Exploration of direct DNA data recording systems within living cells, using optogenetically regulated recombinases and engineered molecular recorders [14-16]. This approach enables the encoding of environmental stimuli, such as light, into DNA, and allows for data processing within biological systems [17,18].

DNA of Things (DoT)

The DoT concept involves embedding digital data into physical objects via DNA tags, enabling independent and offline information storage [19,20]. This has potential for supply chain tracking, anti-counterfeiting, and distributed data storage.

Bio Cybersecurity

As the field matures, the risk of bio-cybersecurity threats increases. This includes the creation of DNA encoded malware. Research into detection and prevention of these threats is increasingly important.

Current Challenges and Future Directions

- Addressing cost and speed limitations in DNA synthesis and sequencing.
- Developing robust and standardized error correction algorithms.
- Improving data retrieval efficiency and scalability.
- Exploring novel encoding and decoding strategies.
- Creating robust bio-cybersecurity protocols.
- Standardization of protocols.

This detailed overview highlights the rapid evolution of DNA digital data storage, emphasizing its potential to revolutionize data management and archiving. Future research efforts focused on addressing current challenges will pave the way for practical and widespread applications.

Intelligent Data Storage: Biomimetic Systems & Genetic Code

The exploration of biomimetic systems, particularly leveraging the genetic code, for intelligent data storage is rapidly gaining momentum. Inspired by the inherent efficiency and robustness of DNA, researchers are actively investigating its unique properties to revolutionize digital information storage [49].

Biomimetic Inspiration: DNA as a Superior Storage Medium

Beyond Binary

Unlike conventional binary storage, DNA utilizes a quaternary system (four nucleotide bases: A, T, C, G), significantly increasing data density. Its three-dimensional structure further enhances storage capacity, offering a substantial advantage over planar electronic storage.

Exceptional Data Density

The human genome, with its 3.2 billion base pairs, exemplifies the extraordinary data density achievable with DNA, far surpassing the capabilities of contemporary hard drives.

Longevity and Stability

DNA's inherent stability enables long-term data preservation, with potential storage durations of hundreds of thousands of years under optimal conditions. This longevity, coupled with relatively low energy consumption, makes DNA a compelling solution for archival storage.

The DNA Data Storage Process: A Five-Step Framework

Encoding

Conversion of digital data into a DNA sequence, addressing challenges such as sequence complexity, GC content optimization, and avoidance of error-prone motifs.

Writing (Synthesis)

Synthesis of the encoded DNA strands, requiring precise control to minimize errors.

Storing

Storage of synthesized DNA, either in vitro (controlled environments) or in vivo (living organisms). In vitro storage requires careful environmental control to prevent degradation. In vivo storage, particularly in microorganisms, offers scalability and potential self-repair mechanisms.

Reading (Sequencing)

Retrieval of the stored DNA sequence using sequencing technologies.

Decoding

Conversion of the sequenced DNA back into digital data.

Encoding and Storage Strategies

Encoding Techniques

Diverse encoding methods, from direct mapping to complex algorithms, are employed to maximize DNA's storage potential.

Storage Modalities

Storage options include in vitro (dry DNA, solutions), devices (microfluidic systems), and in vivo (bacterial cells).

Current Challenges and Future Directions

Technological Limitations

Improving DNA synthesis throughput and sequencing speeds is crucial for practical applications.

Standardization Needs

Standardizing the entire DNA data storage process is essential for interoperability and long-term

accessibility. Metadata integration within the DNA sequence is vital for autonomous data retrieval. Standardization of encoding/decoding algorithms, sample handling, and sequencing protocols is necessary.

Bio Cybersecurity

Addressing potential bio cybersecurity threats is paramount, requiring robust security measures.

Long-Term Accessibility

Developing methods for identifying and interpreting synthetic DNA data in the distant future is crucial. Incorporating unique identifiers, such as rare isotopes or specific DNA sequences, can serve as watermarks.

Medium-Term and Distant-Future Considerations

Medium-Term

Establishing a standardized retrieval mechanism to ensure data accessibility across diverse coding/decoding algorithms.

Distant-Future

Developing robust methods to identify and interpret synthetic DNA data, even if contemporary knowledge is lost. Implementing watermarks to differentiate synthetic from natural DNA.

To better understand and give further context concerning the matters Figure 1 provides an overview visualization concerning the recent computing technicalities in terms of biotechnology retrospectives.

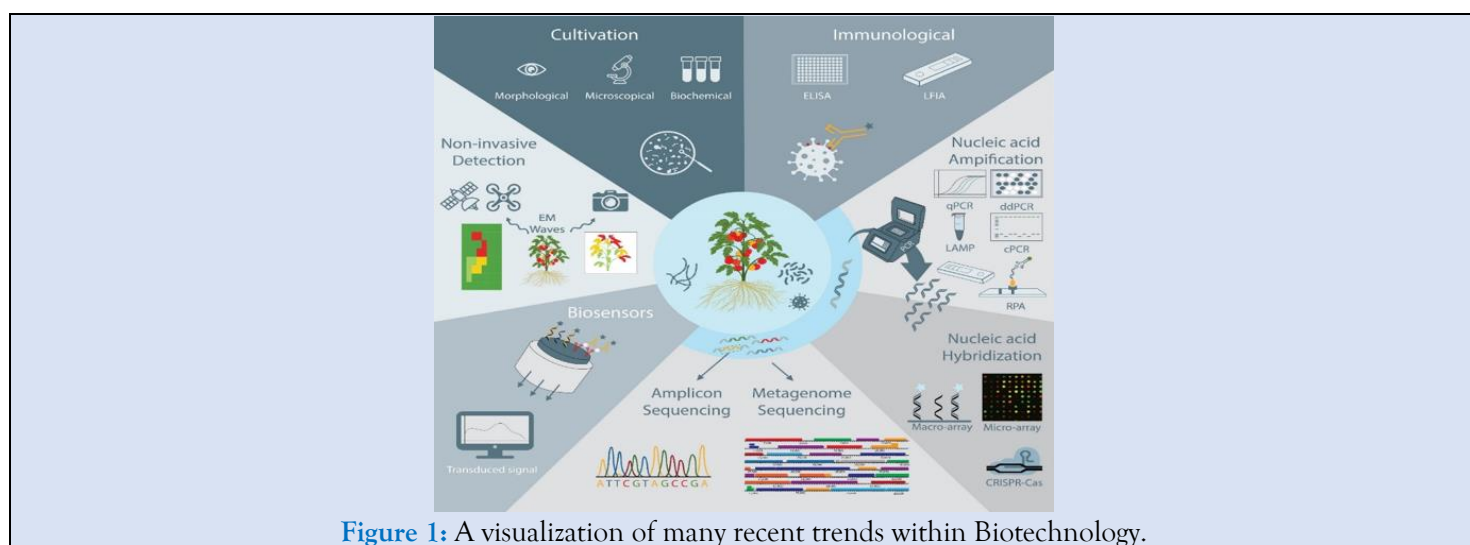


Figure 1: A visualization of many recent trends within Biotechnology.

DNA Data Storage Unlock: Advantages & Challenges

The concept of DNA data storage, initially theorized in the 1990s, gained practical validation in 2012 through the pioneering work of Nick Goldman's team

at the European Bioinformatics Institute. This breakthrough, demonstrating the successful encoding and retrieval of a large text file from synthesized DNA, marked a turning point, propelling DNA data storage into a rapidly evolving field. As of the technical

associated device peripherals for accelerated computing and AI integrations combined towards the human anatomy can hopefully help the future perspectives of the data archival and ease of access along with a probable solution for the management of big data.

Historical Context and Foundational Breakthrough

Early Conceptualization: The idea of leveraging DNA's information storage capacity emerged in the 1990s, laying the groundwork for future research.

Landmark Experiment: Goldman's team's achievement demonstrated the feasibility of encoding and retrieving digital data from synthetic DNA, proving the concept's viability.

Catalyst for Growth: This experiment ignited significant interest, leading to increased research and development efforts by academic institutions and private companies.

Core Principles and Advantages

DNA as a Storage Medium: DNA, the molecule of heredity, possesses inherent characteristics that make it an ideal storage medium. Its high density, long-term stability, and natural error correction mechanisms are key advantages.

High Storage Density: DNA offers an exceptionally high storage density, with a theoretical capacity of up to 1 exabyte per gram, significantly surpassing traditional storage media.

Long-Term Stability: DNA can persist for thousands of years under appropriate storage conditions, making it suitable for archival storage.

This is particularly valuable for preserving scientific data, historical records, and other long-term information.

Inherent Error Correction

While error correction algorithms are still needed, DNA itself has natural error correction properties due to how the molecule is made.

The DNA Data Storage Process

Digital-to-DNA Conversion: Digital data is converted into DNA sequences composed of the four nucleotide bases (A, T, C, G).

DNA Synthesis: The encoded DNA sequences are synthesized using chemical processes.

DNA Storage: The synthesized DNA is stored under appropriate conditions to ensure long-term stability.

DNA Sequencing: The stored DNA is sequenced to retrieve the nucleotide sequence.

DNA-to-Digital Conversion: The sequenced DNA is decoded back into the original digital data.

Technical Challenges and Mitigation Strategies

Cost and Complexity of Synthesis and Sequencing: Historically, the high cost and complexity of DNA synthesis and sequencing have been major obstacles. Recent advancements in these technologies have significantly reduced costs and increased efficiency.

Error Correction: Sequencing errors can introduce inaccuracies, requiring robust error correction mechanisms.

Strategies Include

Redundancy: Storing multiple copies of the data.

Error Correction Algorithms: Using specialized algorithms to detect and correct errors.

Scalability: While density is high, the ability to read and write large amounts of data quickly is still being developed.

Standardization: Standardization of encoding, decoding, and handling procedures is vital for interoperability and long-term accessibility.

Ongoing Advancements and Future Prospects

Technological Improvements: Continued advancements in DNA synthesis and sequencing technologies are driving down costs and improving efficiency.

Algorithm Development: Research into advanced encoding and error correction algorithms is enhancing data accuracy and reliability.

Application Expansion: DNA data storage has potential applications in various fields, including archival storage, data centers, and distributed data storage.

DNA Data Storage: AI Integrations

The intersection of artificial intelligence (AI) and DNA data storage is rapidly transforming the landscape of data management. AI's capabilities are crucial for overcoming the inherent challenges of DNA data storage, enhancing efficiency, and unlocking its full potential. Furthermore, recent breakthroughs, such as the "BacCam" system, demonstrate the transformative power of integrating biological systems with digital technologies. In the upcoming years this innovation could help generate the next level of big data storage device peripherals which can come in the form of integrated systems biology wearables.

AI's Role in Optimizing DNA Data Storage

Sequence Optimization: AI algorithms, particularly machine learning, can analyze and optimize DNA sequences for data storage, considering factors like:

- Data density maximization.
- Error correction code integration.
- Sequencing efficiency.

This optimization ensures that the encoded data is stored efficiently and reliably.

Error Correction Enhancement: AI can significantly improve the accuracy of error correction during DNA sequencing and data retrieval. Machine learning models can identify and correct sequencing errors, enhancing data integrity.

DNA Synthesis Streamlining: AI can optimize DNA synthesis conditions, reducing costs and time. This includes optimizing chemical reactions, predicting synthesis errors, and automating synthesis processes.

Data Compression: AI-driven data compression techniques can minimize the amount of DNA required for storage. This makes DNA data storage more practical and scalable for diverse applications.

The "BacCam" System: A Biological Camera Breakthrough

Innovative Approach: A research team led by Associate Professor Poh Chueh Loo at the National University of Singapore (NUS) developed the "BacCam" system, a biological camera that uses living cells for data storage. This system leverages DNA's exceptional storage capacity, stability, and durability.

Overcoming Traditional Limitations: Unlike traditional DNA storage methods that rely on external DNA synthesis, "BacCam" uses live cells as "data banks." This eliminates the need for external synthesis, making the process more accessible and scalable.

Optogenetic Data Imprinting: The system uses optogenetics to imprint light signals onto DNA within cells, mimicking the process of capturing images on film. This allows for the direct encoding of light-based information into DNA.

Data Organization and Retrieval: Barcoding techniques and machine learning algorithms are used to organize and reconstruct stored images. This mimics the functionality of a digital camera's data capture, storage, and retrieval processes.

Simultaneous Multi-Image Capture: The "BacCam" system can capture and store multiple images simultaneously using different light colors, representing a significant advancement.

Implications and Future Directions

Addressing Data Overload: The "BacCam" system offers a promising solution to the ever-growing data overload problem. It provides a sustainable and high-capacity alternative to traditional data storage methods.

Environmental Impact: By utilizing biological systems, "BacCam" reduces the environmental impact associated with resource-intensive data centers.

Integration of Biological and Digital Systems: Associate Prof. Poh's team emphasizes the integration of biological and digital systems as a key avenue for future exploration and development. Further research will focus on improving the speed and accuracy of the system, and expanding the types of data that can be stored.

AI and Future Development: AI will continue to enhance the "BacCam" system, particularly in image reconstruction and data management.

DNA Tagging Technology: Market Solutions

DNA tagging technology represents a significant advancement in identification, authentication, and tracking, offering solutions that surpass traditional methods. Leveraging DNA's unparalleled storage capacity and stability, this technology is poised to revolutionize various industries.

Fundamentals and Development

Concept and Advantages: DNA tagging offers a robust and secure method for marking objects, particularly those where conventional tagging (UPC barcoding, RFID) is impractical. DNA's high storage density and longevity provide significant advantages.

Design Considerations: Successful DNA tagging requires careful design, including:

- Information type and size.
- Coding methods.
- Storage techniques.
- Data extraction processes.

Historical Challenges and Advancements: Early development faced challenges related to DNA instability and degradation. Protective materials (silica, polymers, gels) and improved decoding methods (PCR, smartphone-based assays, CRISPR-based readout) have mitigated these issues. Portable sequencing technology such as Oxford Nanopore's SmidgION is helping overcome previous obstacles.

DNA Tagging Process and Technologies

Tagging Procedure: DNA is applied to objects through integration or labeling. DNA extraction, amplification, and sequencing are used for tag reading and decoding.

Innovative Systems: "Porcupine" (University of Washington and Microsoft): Portable end-to-end molecular tagging with nanopore signals. DNA ink, DNA-enclosed silica capsules, and adhesive DNA markings.

Commercial Applications: SigNature DNA, Haelixa, SelectaDNA: DNA tagging in RFID devices, labels, and serial numbers for product authentication and anti-counterfeiting.

Holoptica, DNA Technologies: Synthetic DNA tags in inks for brand protection and art authentication, utilizing photoluminescent properties.

Challenges and Limitations

Scalability: Laboratory analysis requirements limit scalability.

Accuracy: Potential for false positives or contamination.

Security: Vulnerability to next-generation sequencing methods necessitates ongoing research.

Applications and Future Potential

Diverse Applications

- Authentication of high-value art and pharmaceuticals.
- Monitoring of weaponry and tracing of valuable minerals.
- Data security and encryption in sensitive sectors (defense, healthcare).
- Forensic science, animal tracking, and industrial process management.

Technological Advancements

- Improved DNA synthesis and encoding technologies.
- Portable sequencing technologies for everyday integration.

Emerging Concepts

DNA-of-Things: Linking physical objects to unique DNA identifiers.

Metaverse Integration: Linking physical objects to digital representations (NFTs).

Enhanced Security: Combining DNA tagging with cryptography for robust security.

Forensic Science and Security

- DNA sprays for criminal tracking.
- Animal tagging without intrusive devices.
- Military and intelligence applications, including quantum computing deterrence.
- Cybersecurity defense through DNA labeling and encryption.

Food and Pharmaceutical Safety

- Anti-counterfeiting measures to combat counterfeit medication.
- Enhanced safety and security in production.
- Addressing the rising threat of counterfeit drugs, particularly highlighted during the COVID-19 pandemic.

Interdisciplinary Collaborations

- The interdisciplinary nature of DNA tagging necessitates collaboration among experts from various fields.
- Continued research and development are crucial for addressing challenges and expanding applications.

To better understand and give further context concerning the matters Figure 2 provides an overview visualization concerning the DNA tagging computing technicalities in terms of biotechnology retrospectives.

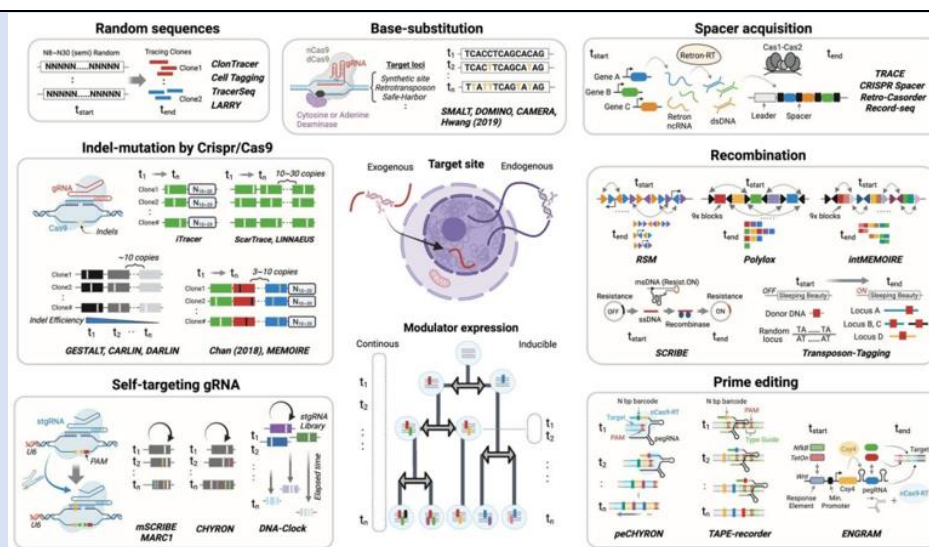


Figure 2: DNA Tagging Technology perspectives.

Results and Findings

This research has synthesized and analyzed the current state of DNA data storage, revealing key findings regarding its potential and the challenges hindering its widespread adoption. To provide a better understanding relating to the matters of perspectives Figure 3-5 provides the associated research results along with the findings concerning these investigative explorations.

Current Limitations and Benchmarks

Writing Speed and Capacity: DNA data storage currently faces significant limitations in writing speed, with the current record at approximately 200 MB. Single synthesis runs require around 24 hours, posing a substantial bottleneck for practical applications. This significant limitation makes DNA data storage currently uncompetitive with traditional electronic data storage.

Cost Constraints: The cost of DNA synthesis and sequencing remains a major barrier, inhibiting the scalability and accessibility of the technology.

Advancements in Encoding, Writing, and Reading Processes

Encoding Scheme Improvements: Ongoing efforts are focused on developing more efficient encoding schemes to increase data density and reduce synthesis complexity.

Chemical and Enzymatic Writing Processes: Emerging chemical and enzymatic processes are driving down the costs and time associated with DNA synthesis.

These advancements are crucial for both sequence-based and structure-based DNA storage approaches.

DNA Sequencing Enhancements

Advancements in DNA sequencing techniques are critical for improving data readout speed and accuracy. Exploration of new chemistries, such as unnatural nucleotides and molecules modulating DNA structure, is expanding the parameter space for both sequence- and structure-based storage.

Readout Technologies

The integration of solid-state nanopores and optical techniques in readout processes offers the potential for enzyme-free, high-accuracy, and high-speed data retrieval. DNA nanotechnology assembly procedures are showing promise.

Computational Methods

Computational methods are being increasingly utilized to improve all aspects of DNA data storage, from encoding, to error correction, and to decoding.

Archival Storage and Long-Term Stability

Archival Application: Archival storage is identified as a primary application for DNA data storage, leveraging its long-term stability.

Longevity Research: Further research is required to evaluate the longevity of noncovalently assembled DNA nanostructures and their readability after prolonged storage.

Dynamic DNA Databases: Implementing dynamic properties within DNA databases, such as data erasure, rewriting, and updates, is crucial for enhancing practical viability and efficiency.

Multidisciplinary Collaboration and Holistic Approach

Collaborative Efforts: A concerted multidisciplinary effort is essential to propel the field forward, requiring collaboration among researchers from diverse areas. This includes expertise in chemical techniques, instrumentation, characterization methods, and automated analysis tools.

Holistic Bottom-Up Design: A holistic bottom-up approach is crucial for designing the entire process, from data encoding to decoding. This requires collaborative efforts across various scientific disciplines, including mathematics and polymer chemistry.

Key Findings

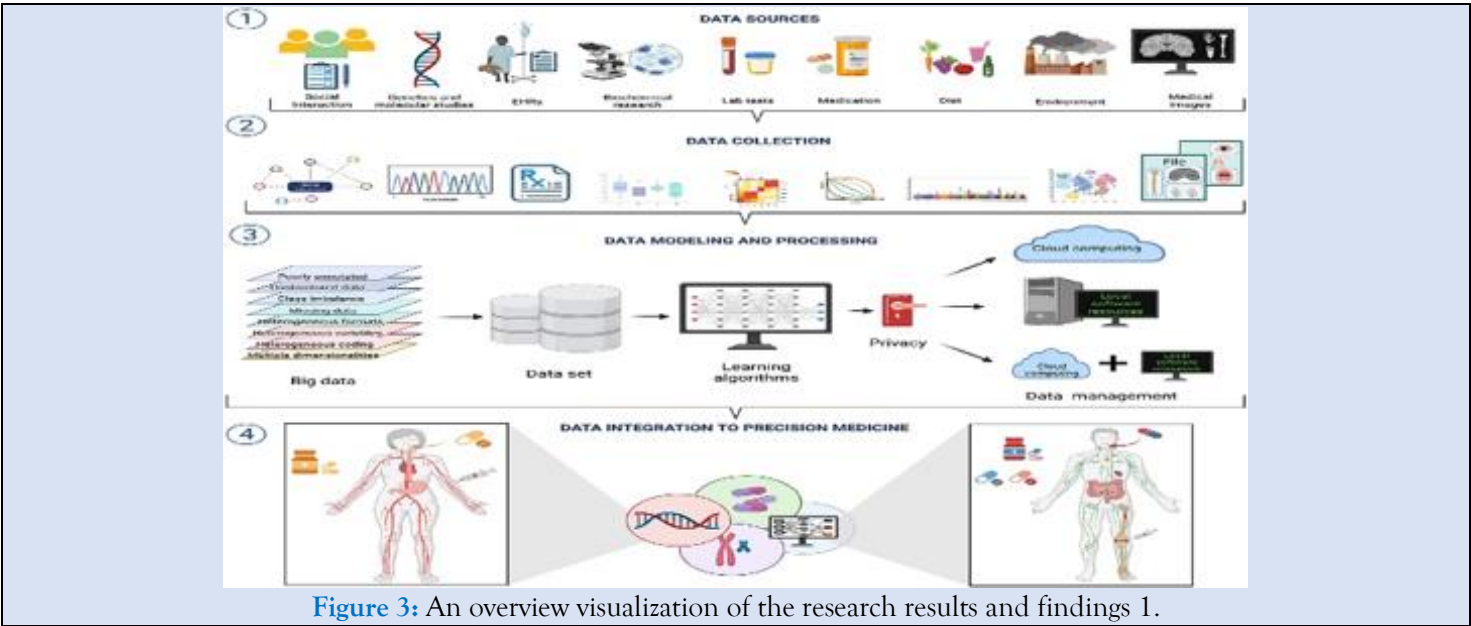
While DNA data storage holds immense potential for high-density information storage, significant limitations in writing speed and cost currently hinder its competitiveness.

Advancements in encoding schemes, writing and reading processes, and storage procedures are driving progress in the field.

Archival storage is a promising application for DNA data storage, but further research is needed to evaluate long-term stability and implement dynamic database properties.

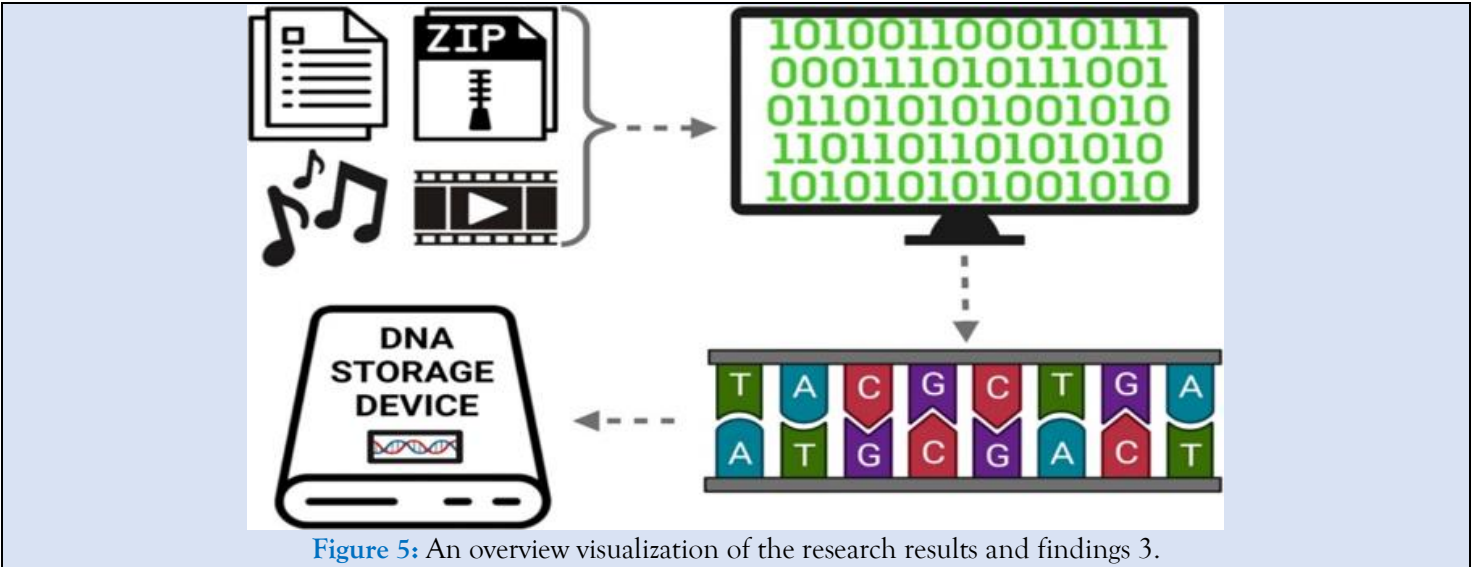
Multidisciplinary collaboration and a holistic approach are essential for the successful development and deployment of DNA data storage systems.

The use of computational methods is becoming an integral part of the DNA data storage process.



message	r	a	m	y
ASCII	114	97	109	121
binary sequence	01110010	01100001	01101101	01111001
DNA sequence	CTAG	CGAC	CGTC	CTGC

Figure 4: An overview visualization of the research results and findings 2.



There are various types of company and organizations who are providing technology features within specific markets. This research investigative explorations have been resourced from various types of domains and associated platform of several data resources and their availability of many materials which are referenced within Table 1 for further information.

Table 1: An overview of all the associated data materials and resources.

Company Name, Country of Origin and Launch Date	Main Features of The Technology	Selected Markets	References
Applied DNA Sciences, United States, 1983	Botanical DNA fragments, detection by PCR and CE, an encapsulation system	Product authentication, supply chain traceability, brand protection, anti-counterfeiting, textiles, pharmaceuticals, etc.	[21-26]
Haelixa, Switzerland, 2016	Synthetic DNA tags, detection by PCR, DNA enclosed in silica	Product authentication, supply chain traceability, intellectual property protection, etc.	[26-30]
Selectamark Security Systems (SelectaDNA), United Kingdom, 1986	Laboratory analysis of DNA for owner identification if microdots are absent (DNA serves as an alternative authentication solution)	Asset protection and recovery, securing high-value items, art and jewelry authentication, IT equipment and vehicle security, forensic applications, theft prevention and deterrence, etc.	[31-33]
TraceTag (CypherMark), United Kingdom / Norway, 2001	Synthetic DNA with unique primers, authorized access to primer sequences, detection using qPCR	Brand safeguarding, industrial applications, cash security, security of documentation, oil and fuel tracking, anti-counterfeiting measures, etc.	[34-36]
Holoptica, United States, 2012	Synthetic DNA tags (100 nucleotides), integration with inkjet cartridges	Artwork, documents and assets protection, verifying product authenticity, food tracking, etc.	[37-39]
DNA Technology, United States, 1993	DNA-laced ink, combination of DNA synthetic segments and optical taggants	Memorabilia and collectibles, limited edition artwork, pharmaceuticals, apparel and luxury goods, health and beauty industry, etc.	[40,41]
Tagsmart, United Kingdom, 2015	Synthetic DNA tags, secure Certificate of Authenticity	Artwork, securing collectibles, verifying paper documents, book manufacturing, etc.	[42,43]
DNA Guardian, Australia, 2007	UV-detectable stain, detection using pyrosequencing	Asset marking, crime prevention, artwork protection, theft deterrence, etc.	[44,45]
Aanika Biosciences, United States, 2018	Genetically modified <i>Bacillus subtilis</i> as an encapsulation system for DNA tag	Agriculture and food production, textiles, etc.	[46-48]

Discussions and Future Directions

The burgeoning field of DNA data storage is demonstrating significant potential, as evidenced by recent market initiatives and ongoing technological advancements. This section delves into the discussions surrounding its current state, future directions, and strategic implications.

Market Trends and Applications

Emerging Market Initiatives: Collaborations like the Netflix-Twist Bioscience partnership, encoding the "Biohackers" series into DNA, highlight the growing interest in DNA data storage.

Current Application Focus: DNA data storage is currently most viable for cold and glacial storage due to limitations in writing, reading, and decoding speeds for warm or hot storage applications.

Anticipated Advancements: Advancements in writing platforms are expected to enable massive parallel DNA synthesis, achieving gigabytes on a single chip in the coming years and terabytes within the next decade. The inherent capacity and stability of DNA offer significant advantages over traditional storage mediums.

Challenges to Overcome: Enhancing the speed and length of synthetic DNA writing is crucial for broader adoption. Developing efficient file location methods

and reagent/nucleotide reusability is essential for hot storage applications. Enzymatic synthesis holds promise for addressing writing speed limitations.

Technological and Economic Considerations

Cost Reduction: Increased adoption is expected to drive down costs, making DNA data storage more affordable. Energy-efficient storage systems offer long-term cost benefits compared to traditional data servers.

Industry Collaboration: Active participation from various companies across the DNA data storage pipeline, from CODEC development to synthesis and sequencing, is crucial for widespread access within the next decade.

Accessibility: The goal is to enable the general public and businesses to store diverse binary data into DNA, ensuring long-term preservation.

Strategic Implications and Investment

Visionary Solution: DNA data storage presents a visionary solution for archiving, research, finance, and technology sectors.

Pragmatic Approach: While enthusiasm is warranted, a pragmatic approach is essential, considering both potential and ongoing research.

Strategic Investment: Strategic investment in DNA data storage offers future-proof capabilities,

surpassing current electronic and magnetic storage limitations. DNA's stability and longevity make it a prudent choice for long-term data storage.

Diversification and Leadership: Incorporating DNA data storage diversifies data management strategies, mitigating risks. Early adopters gain a competitive edge, setting industry standards in efficiency and innovation.

Wide-Ranging Applications: DNA data storage is invaluable for preserving historical records, managing research data, and ensuring security in finance and technology sectors.

Collaborative Innovation: Investment fosters partnerships between biotechnologists, data scientists, and IT professionals.

Environmental Sustainability: DNA data storage aligns with sustainability goals, offering a green technology solution.

Phased Implementation: A phased investment approach, starting with pilot projects and research initiatives, allows for effective understanding and leverage of the technology.

Future Directions and Regulatory Considerations

Regulatory Mapping: Regulatory directions need to be mapped towards the technical underpinnings, advantages, potential applications, challenges, and future prospects of DNA data storage.

Balancing Enthusiasm and Pragmatism: Implementations must balance enthusiasm with pragmatic considerations, addressing current challenges such as high costs, slower access speeds, and technical complexities.

Continuous Evolution: DNA data storage technology represents a strategic move towards innovative, sustainable data management, positioned at the cutting edge of data-driven advancements.

The Future: Organizations adopting this technology are prepared to leverage its immense potential as it evolves into a pivotal solution for managing the ever-growing digital landscape.

Conclusion

DNA data storage emerges as a transformative technology, poised to revolutionize long-term data preservation. Its inherent security, exceptional density, and remarkable longevity present a compelling alternative to traditional storage methods. While current technical and economic hurdles persist, the relentless pace of research and development signals a promising future.

The convergence of advancements in DNA synthesis, sequencing, and artificial intelligence is poised to incrementally bridge the gap between theoretical potential and practical application, making DNA data storage increasingly viable and cost-effective. Ultimately, this technology offers a robust solution for addressing the escalating demand for secure and enduring data archiving. The sheer volume of data storable within DNA is contingent upon a confluence of factors, including the length of DNA sequences, the efficiency of encoding algorithms, and the precision of sequencing and retrieval processes.

While the theoretical storage capacity of DNA, reaching an astounding 1 exabyte per gram, remains a beacon of its potential, practical constraints imposed by contemporary DNA synthesis and sequencing technologies, coupled with associated costs and complexities, currently limit its realization.

Nevertheless, the ongoing synergy between research endeavors and technological innovation, particularly the integration of AI, holds the promise of significantly expanding the practical data storage capacity of DNA in the foreseeable future. It is imperative to acknowledge that DNA data storage, while possessing unparalleled archival capabilities, may not be universally applicable across all data storage scenarios. Its strengths lie in the realm of large-scale, long-term data archives, such as scientific datasets and historical records, where data integrity and longevity are paramount. Conversely, applications demanding high-speed data retrieval or frequent data modification may find DNA data storage less suitable.

This nuanced understanding underscores the importance of continued research and development efforts aimed at addressing the current limitations and maximizing the technology's potential across a diverse spectrum of data storage contexts. As the field matures, a strategic approach that leverages DNA data storage for its unique strengths, while acknowledging its current limitations, will be crucial for its successful integration into the broader data storage ecosystem.

Supplementary Information

The various original data sources some of which are not all publicly available, because they contain various types of private information. The available platform provided data sources that support the exploration findings and information of the research investigations are referenced where appropriate.

Acknowledgments

The authors would like to acknowledge and thank the GOOGLE Deep Mind Research with its associated pre-prints access platforms. This research exploration was investigated under the platform provided by GOOGLE Deep Mind which is under the support of the GOOGLE Research and the GOOGLE Research Publications within the GOOGLE Gemini platform. Using their provided platform of datasets and database associated files with digital software layouts consisting of free web access to a large collection of recorded models that are found within research access and its related open-source software distributions which is the implementation for the proposed research exploration that was undergone and set in motion. There are many data sources some of which are resourced and retrieved from a wide variety of GOOGLE service domains as well. All the data sources which have been included and retrieved for this research are identified, mentioned and referenced where appropriate.

Declarations

Funding

No Funding was provided for the conduction concerning this research.

Conflict of Interest/Competing Interests

There is no Conflict of Interest or any type of Competing Interests for this research.

Ethics Approval

The authors declare no competing interests for this research.

Consent to Participate

The authors have read, approved the manuscript and have agreed to its publication.

Consent for Publication

The authors have read, approved the manuscript and have agreed to its publication.

Availability of Data and Materials

The various original data sources some of which are not all publicly available, because they contain various types of private information. The available platform provided data sources that support the exploration findings and information of the research investigations are referenced where appropriate.

Code Availability

Mentioned in details within the Acknowledgements section.

Authors' Contributions

Described in details within the Acknowledgements section.

References

1. Ceze, L., Nivala, J., Strauss, K. (2019). Molecular Digital Data Storage Using DNA. *Nature Reviews Genetics*, 20(8):456-466.
2. Akram, F., Haq, I. U., Ali, H., Laghari, A. T. (2018). Trends to Store Digital Data in DNA: An Overview. *Molecular Biology Reports*, 45(5):1479-1490.
3. Panda D, Molla KA, Baig MJ, Swain A, Behera D, et al. (2018). DNA As a Digital Information Storage Device: Hope or Hype? *3 Biotech*. 8(5):239.
4. Shankland, S. (2019). Startup Packs All 16GB of Wikipedia onto DNA Strands to Demonstrate New Storage Tech.
5. Roquet, N., Bhatia, S. P., Flickinger, S. A., Mihm, S., Norsworthy, M. W., et al. (2021). DNA-Based Data Storage via Combinatorial Assembly. *bioRxiv*.
6. Smith GC, Fiddes CC, Hawkins JP, Cox JP. (2003). Some Possible Codes for Encrypting Data in DNA. *Biotechnology Letters*. 25(14):1125-1130.
7. Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, et al. (2013). Towards Practical, High-Capacity, Low-Maintenance Information Storage in Synthesized DNA. *Nature*. 494(7435):77-80.
8. Lee HH, Kalhor R, Goela N, Bolot J, Church GM. (2019). Terminator-Free Template-Independent Enzymatic DNA Synthesis for Digital Information Storage. *Nature Communications*. 10(1):2383.
9. Kalhor R, Kalhor K, Mejia L, Leeper K, Graveline A, et al. (2018). Developmental Barcoding of Whole Mouse via Homing CRISPR. *Science*. 361:(6405).
10. Lim CK, Yeoh JW, Kunartama AA, Yew WS, Poh CL. (2023). A Biological Camera That Captures and Stores Images Directly Into DNA. *Nature Communications*. 14(1):3921.
11. Rebrova IM, Rebrova OY. (2020). Storage Devices Based on Artificial DNA: The Birth of An Idea and The First Publications. *Voprosy Istorii Estestvoznaniia i Tekhniki*. 41(4):666-676.
12. Extance A. (2016). How DNA Could Store All the World's Data. *Nature*. 537(7618):2224.
13. Limbachiya D, Dhameliya V, Khakhar M, Gupta MK. (2016). On Optimal Family of Codes for

- Archival DNA Storage. 2015 Seventh International Workshop on Signal Design and Its Applications in Communications (IWSDA). 123-127.
14. Organick L, Ang SD, Chen YJ, Lopez R, Yekhanin S, et al. (2018). Random Access in Large-Scale DNA Data Storage. *Nature Biotechnology*. 36(3):242-248.
15. Patel P. (2018). DNA Data Storage Gets Random Access. *IEEE Spectrum: Technology, Engineering, and Science News*.
16. Langston J. (2019). With a Hello, Microsoft and UW Demonstrate First Fully Automated DNA Data Storage. *Innovation Stories*.
17. Appuswamy R, Le Brigand K, Barbry P, Antonini M, Madderson O, et al. (2019). OligoArchive: Using DNA in The DBMS Storage Hierarchy. *Conference on Innovative Data Systems Research (CIDR)*.
18. Tabatabaei SK, Wang B, Athreya NB, Enghiad B, Hernandez AG, et al. (2020). DNA Punch Cards for Storing Data on Native DNA Sequences via Enzymatic Nicking. *Nature Communications*. 11(1):1742.
19. Molteni M. (2019). These Plastic Bunnies Got a DNA Upgrade. Next up, the World? *Wired*.
20. Hotz RL. (2019). Scientists Store Data in Synthetic DNA Embedded in a Plastic Bunny. *Wall Street Journal*.
21. Han, S., Guo, X., Jin, X., Shen, X. (1998). Methode a Technique Analytique de Code D'acide Nucleique Utilisee Dans des Etiquettes Infalsifiables. *WO patent WO1998006084A1*.
22. Pierre, A., Bonnet, J., Vekris, A., Portier, J. (2001). Encapsulation of Deoxyribonucleic Acid Molecules in Silica and Hybrid Organic-Silica Gels. *J. Mater. Sci. Mater. Med.* 12:51-55.
23. Gootenberg, J. S. et al. (2017). Nucleic acid Detection with CRISPR-Cas13a/C2c2. *Science*, 356:438-442.
24. Qian, J. et al. (2020). Barcoded Microbial System for High-Resolution Object Provenance. *Science*, 368:1135-1140.
25. SmidgION. Oxford Nanopore Technologies.
26. Hashiyada, M. (2004). Development of Biometric DNA Ink for Authentication Security. *Tohoku J. Exp. Med.* 204:109-117.
27. Sharief, S. A., Chahal, P. Alocilja, E. (2021). Application of DNA Sequences in Anti-Counterfeiting: Current Progress and Challenges. *Int. J. Pharm.* 602:120580.
28. Altamimi, M. J. et al. (2019). Anti-Counterfeiting DNA Molecular Tagging of Pharmaceutical Excipients: An Evaluation of Lactose Containing Tablets. *Int. J. Pharm.* 571:118656.
29. Puddu, M., Paunescu, D., Stark, W. J. Grass, R. N. (2014). Magnetically Recoverable, Thermostable, Hydrophobic DNA/Silica Encapsulates and Their Application as Invisible Oil Tags. *ACS Nano*, 8:2677-2685.
30. Hogan, M. E., Jung, L., Richter, N., Sun, Y. Szczepanik, M. (2018). Nucleic Acid Coated Submicron Particles for Authentication. *WO patent WO2018156352A1*.
31. Ponnuswamy, N. et al. (2017). Oligolysine-Based Coating Protects DNA Nanostructures from Low-Salt Denaturation and Nuclease Degradation. *Nat. Commun.* 8:15654.
32. Brown, J., Reichert, B. (2019). Compositions for Use in Security Marking. *US patent US20120135413A1*.
33. Sleat, R., Van Lint, G. (2014). Marking Apparatus for Nucleic Acid Marking of Items. *US patent US8783194B2*.
34. Dollinger, G. D. (1990). Methods for Tagging and Tracing Materials with Nucleic Acids. *WO patent WO1990014441A1*.
35. Berk, K. L. et al. (2021). Rapid Visual Authentication Based on DNA Strand Displacement. *ACS Appl. Mater. Interfaces*, 13:19476-19486.
36. Jung, L., Hayward, J. A., Liang, M. B. Berrada, A. (2018). DNA Marking of Previously Undistinguished Items for Traceability. *US patent US9919512B2*.
37. Crowley, J., Rademan, K., Tan, E. (2022). The Textile Tracer Assessment. *Fashion for Good Reports*.
38. Securing Industry. (2018). Inkjet Codes with sigNature® Molecular Tags for Supply Chain Traceability and Brand Protection.
39. Mead, B. K. et al. (2014). The Role of DNA-Stained Currency in Gang Robberies Within the United Kingdom. *J. Forensic Sci.* 59:264-267.
40. Sniderman, D. (2017). Are Synthetic Chemical Markers Needed in Lubricants? *Tribol. Lubr. Technol.* 73:38-46.
41. Ullrich, C., Luescher, A. M., Koch, J., Grass, R. N. Sax, H. (2022). Silica Nanoparticles with Encapsulated DNA (SPED) to Trace the Spread

- of Pathogens in Healthcare. *Antimicrob. Resist. Infect. Control*, 11:4.
42. DENIMSANDJEANS. (2022). How Soorty Is Leveraging Haelixa DNA to Get Brands into Traceability. *Denimandjeans | Global trends, news and reports | Worldwide*.
43. Paunescu, D., Puddu, M., Soellner, J. O. B., Stoessel, P. R., Grass, R. N. (2013). Reversible DNA Encapsulation in Silica to Produce ROS-Resistant and Heat-Resistant Synthetic DNA 'fossils', *Nat. Protoc.* 8:2440-2448.
44. SelectaDNA. (2023). SelectaDNA Forensic Coding.
45. Pakendorf, M. (2019). Combinatorial DNA Barcoding for Security Applications. *Doctoral Dissertation, University of Leicester, Leicester*.
46. Bonnell, A. (2010). Covert Tagging. *Pira International Ltd, United Kingdom*.
47. Brand Integrity International. (2023). DNA smartMarks. *BrandProtection*.
48. Holoptica. (2023). DNA Technology.
49. Akhtar ZB. (2025). Exploring the Role of Molecular Engineering in Regenerative Medicine, Generis Publishing.

Cite this article: Akhtar ZB, Rawol AT. (2026). DNA Data Storage: Multidisciplinary Approaches to Long-Term Digital Preservation Towards Advancements, Challenges, Future Prospects, *International Clinical Case Reports and Reviews*, BioRes Scientia Publishers. 4(1):1-14. DOI: 10.59657/2993-0855.brs.26.045

Copyright: © 2025 Zarif Bin Akhtar, this is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Article History: Received: March 31, 2025 | Accepted: July 24, 2026 | Published: August 03, 2026