

Harnessing Bacteria for Next-Generation Data Storage Technologies – A Review

Bhuvi^{1,*}, Nitish Kumar Raghav²

Abstract

The exponential increase in global digital information has created a pressing need for storage technologies that are more durable, compact, and sustainable than conventional electronic media. While hard drives, solid-state drives, and cloud-based systems have transformed information management, they face severe limitations related to storage density, energy consumption, maintenance costs, and long-term preservation. Researchers have, therefore, begun exploring biological systems as alternative information storage platforms. Among these, bacteria have emerged as particularly promising candidates because of their ability to naturally store, replicate, and transfer genetic information through DNA. This review examines the conceptual framework of bacterial data storage, the fundamental scientific principles underlying DNA-based information encoding, recent breakthroughs in synthetic biology, and the advantages and limitations of living storage systems. Furthermore, we evaluate the expanding architecture of molecular networks, error-correction mechanisms, and the prospective deployment of engineered microorganisms for long-term archival storage. By integrating state-of-the-art computational algorithms with molecular biology, synthetic data sequences can be embedded into plasmids or genomic loci without disrupting cellular viability. Ultimately, this article highlights how bacteria may contribute to the development of next-generation archival storage technologies, bridging the gap between biological efficiency and digital infrastructure. Through a systematic evaluation of current bottlenecks, such as retrieval latency and mutation rates, we outline the trajectory required to bring this biological computing paradigm to commercial scalability.

Keywords: Bacteria, DNA, biological storage, preservation, plasmids, bioinformatics

INTRODUCTION

Human civilization is generating digital information at an unprecedented rate. Every photograph uploaded to social media, every scientific experiment, every financial transaction, and every healthcare record contributes to a continuously expanding digital universe. This data deluge is driven by the rise of the Internet of Things (IoT), artificial intelligence, high-throughput scientific sequencing, and pervasive mobile computing. Experts estimate that global data production is increasing so rapidly that traditional storage technologies may struggle to keep pace in the coming decades. The global datasphere, which reached dozens of zettabytes in the early 2020s, is projected to grow exponentially, overwhelming our industrial capacity to manufacture conventional silicon-based memory and magnetic media [1].

*Author for Correspondence

Bhuvi

E-mail: nitish.raghav@gmail.com

¹Student, St. Joseph Boy's School, Defence Colony, Jalandhar, Punjab, India

²Assistant Professor, Department of Pharmaceutical Sciences, DAV University, Jalandhar, Punjab, India

Received Date: June 09, 2026

Accepted Date: June 12, 2026

Published Date: June 20, 2026

Citation: Bhuvi, Nitish Kumar Raghav. Harnessing Bacteria for Next-Generation Data Storage Technologies – A Review. International Journal of Molecular Biotechnological Research. 2026; 4(2): 1–9p.

Conventional storage devices – such as magnetic tapes, optical discs, hard disk drives (HDDs), and flash memories (SSDs) – have served society well over the past half-century. However, these technologies are associated with fundamental physical limitations including hardware degradation, high energy requirements, periodic maintenance, and restricted lifespans. Data centers currently

consume an enormous amount of global electricity, a substantial portion of which is dedicated solely to thermal management and preventing mechanical failure. The carbon footprint of these immense facilities poses a significant challenge to global sustainability goals. Moreover, electronic media require periodic data migration every 3 to 10 years to prevent data corruption caused by magnetic degradation or physical wear – a process that is both costly and highly susceptible to data loss. As a result, researchers are increasingly searching for storage media capable of preserving information for decades or even centuries while occupying minimal physical space and demanding zero power during preservation [2].

DNA has attracted significant attention because it is nature's own information storage molecule. For billions of years, living organisms have relied on DNA to store and transmit complex genetic information across thousands of generations without losing essential behavioral or physiological data. Remarkably, DNA possesses an information density far greater than that of current electronic devices. Scientists have demonstrated that enormous quantities of information can be stored within microscopic amounts of DNA; theoretically, a single gram of dry DNA can hold up to 215 petabytes of data. To put this into perspective, all the data currently accessible on the global internet could theoretically be stored in a biological matrix roughly the size of a shoebox. Furthermore, when kept in an environment isolated from water, light, and oxygen, DNA can remain stable for millennia, as evidenced by the successful sequencing of genomes from ancient woolly mammoths and archaic hominids that lived hundreds of thousands of years ago [3].

Bacteria represent an especially attractive biological storage platform because they naturally maintain DNA, replicate it with extreme fidelity during cell division, and transfer it through mechanisms such as horizontal gene transfer and conjugation. Unlike synthetic, cell-free DNA storage systems – which require expensive in vitro amplification and are highly susceptible to physical degradation, hydrolysis, and oxidative damage over time – living bacteria encapsulate and safeguard the encoded sequences within protective cellular walls. They provide autonomous maintenance, error repair via endogenous DNA polymerases, and cost-effective replication through simple metabolic growth. These unique biological properties have inspired researchers to investigate whether bacteria can function as living memory devices capable of preserving digital information across unprecedented timescales, effectively creating living hard drives that leverage evolutionary biology for digital permanence [4].

AIM OF THE STUDY

To rigorously evaluate the potential of bacteria as biological systems for the ultra-high-density, long-term storage and preservation of digital information, assessing both the underlying molecular mechanisms and the computational architectures required to bridge binary code with biological mediums.

OBJECTIVES OF THE STUDY

- To understand the fundamental scientific principles of bacterial data storage and molecular encoding.
- To examine the computational and chemical methods used for translating and encoding binary digital information into quaternary DNA structures.
- To analyze the specific role of plasmids, horizontal gene transfer, and bacterial genetics in establishing scalable information storage networks.
- To systematically review recent advancements (2024–2026) in “in vivo” DNA-based storage technologies, including CRISPR logging and AI-driven retrieval models.
- To critically evaluate future opportunities, biophysical limitations, and ethical challenges associated with biological data storage.

MATERIALS AND METHODS

This review employs a systematic scoping methodology to ensure a rigorous and replicable synthesis of peer-reviewed scientific literature related to microbiology, molecular genetics, synthetic biology, bioinformatics, DNA data storage, and genome engineering. To capture both foundational paradigms and cutting-edge innovations, the database selection criteria prioritized high-impact indices,

comprehensively analyzing recent publications from leading scientific databases such as PubMed, Scopus, and IEEE Xplore. Specific attention was given to seminal papers from journals such as Nature, Science, Nucleic Acids Research, and the IEEE Transactions on NanoBioscience [5].

The literature search was strategically constrained to papers published between 2015 and 2026, focusing on keywords including “in vivo DNA storage”, “bacterial memory systems”, “CRISPR data logging”, “biological cryptography”, and “enzymatic DNA synthesis”. This specific temporal scope was intentionally selected to encompass the critical evolution from early theoretical propositions of molecular storage to the latest empirical demonstrations of in vivo data manipulation. Studies that purely focused on “in vitro” (cell-free) DNA storage without application or direct relevance to living biological hosts were excluded to maintain the focus on bacterial systems. The gathered data were thematically synthesized to identify current developments, technological achievements, and critical bottlenecks in biological data storage research, moving beyond a mere chronological summary to provide a deeply integrated analysis of the consensus and remaining gaps within the field [6].

BIOLOGICAL BASIS OF INFORMATION STORAGE

DNA consists of four nucleotide bases: adenine (A), thymine (T), guanine (G), and cytosine (C). The arrangement of these bases forms the genetic code responsible for controlling biological functions, protein synthesis, and metabolic pathways in living organisms. The double-helix architecture, stabilized by hydrogen bonding between complementary bases (A-T and G-C) and hydrophobic stacking interactions, gives DNA its structural resilience. From an information science perspective, DNA can be viewed as a natural storage medium in which nucleotide sequences represent encoded information, effectively functioning as a quaternary (four-base) digital drive. Consequently, a consensus has emerged within the synthetic biology community that leveraging these natural architectures offers an unprecedented pathway to overcoming the physical density limits of conventional silicon storage [7].

Bacteria contain both chromosomal DNA, which holds the core genes necessary for cell survival, and plasmids, which are small, circular, double-stranded DNA molecules capable of independent replication. Plasmids are found in the bacterial cytoplasm and are naturally distributed during cell fission. Because plasmids can be easily extracted, modified through recombinant DNA techniques, and re-introduced into cells, they provide a highly versatile and convenient platform for introducing artificial, information-bearing DNA sequences into bacterial hosts without disrupting vital physiological mechanisms. The cellular machinery of the bacterium, including its native polymerases, helicases, and topoisomerases, essentially act as biological read/write heads that process the artificial DNA exactly as they would natural genetic material [8].

The extraordinary storage capacity of DNA is one of its most remarkable characteristics. Theoretically, a very small quantity of DNA can store information equivalent to vast electronic databases. While an industrial data center requires thousands of square meters of space and megawatts of cooling power, an equivalent volume of data could be housed in a few milligrams of bacterial culture. This property has stimulated significant interest in developing DNA-based information systems that could replace traditional magnetic archives, introducing a paradigm shift where data density is measured at the molecular scale rather than the nanometer scale of silicon gates. In these bacterial systems, the spatial organization of the nucleoid and the thermodynamic stability of supercoiled plasmids provide a naturally optimized physical environment that protects the encoded data from spontaneous degradation, ensuring structural integrity far superior to what can be achieved in a sterile, cell-free synthetic matrix [9].

PRINCIPLES OF BACTERIAL DATA STORAGE

Transitioning from molecular biology to digital encoding, the translation of binary structures into biological mediums presents distinct computational challenges. The process begins by converting digital information into binary code, the native language of electronic computers composed of 0s and 1s. Specialized computational algorithms then translate this binary data into nucleotide sequences. For example, a basic mapping might allocate “00” to A, “01” to C, “10” to G, and “11” to T [10].

Once the sequence is mathematically optimized to avoid biological incompatibility, these synthetic DNA fragments are produced through chemical (phosphoramidite) or advanced enzymatic DNA synthesis technologies. The precise stoichiometric requirements of these synthesis reactions are critical to ensuring the structural fidelity of the data strands. Following synthesis, the artificial DNA sequences are inserted into bacterial plasmids using restriction enzymes and ligases or directly integrated into the bacterial chromosome via homologous recombination or specialized transposons [11].

Once incorporated into bacterial cells, the encoded information becomes an integral part of the organism's genetic material. As bacteria grow and divide, they replicate the stored DNA using their native high-fidelity polymerases, effectively generating multiple copies of the information across billions of daughter cells within hours. Crucially, this natural replication process provides an inherent, fully automated backup mechanism that requires no external manual copying or technical intervention, ensuring data redundancy on an astronomical scale [12].

Data retrieval involves isolating the recombinant plasmid or genomic DNA from the bacterial cells and sequencing the encoded regions using next-generation sequencing (NGS) platforms or third-generation nanopore sequencing. Bioinformatics tools subsequently process the raw sequencing reads, filter out biological noise, and convert the nucleotide sequences back into their original digital format. Robust error-correction systems are incorporated into the software pipeline to ensure accurate, bit-perfect recovery of information despite any biochemical variations that may have occurred during the life cycle of the bacteria [13].

ENCODING AND DECODING STRATEGIES

One of the most important aspects of DNA storage involves the design of efficient encoding algorithms. Early approaches used simple, direct conversions between binary digits and nucleotide bases, which frequently resulted in problematic sequences that living cells struggled to process. For example, long homopolymer runs (such as AAAAAA or GGGGGG) are notorious for causing slip errors during both chemical synthesis and sequencing. In living organisms, these repetitive structures can cause DNA polymerases to stutter, leading to insertion or deletion mutations that catastrophically corrupt the encoded file. Furthermore, extreme Guanine–Cytosine (GC) content imbalance can cause structural instability in the DNA double helix due to the stronger triple hydrogen bonds between G and C compared to the double bonds between A and T, making it difficult to amplify or sequence accurately [14].

To circumvent these biological barriers, recent literature highlights a paradigm shift toward constrained coding schemes and advanced error-correction codes (ECC), like those utilized in satellite telecommunications and optical storage. Modern advanced systems incorporate sophisticated compression methods and mapping algorithms that minimize repetitive nucleotide patterns and maintain a balanced GC ratio near 50%. Researchers have successfully encoded text documents, historical images, high-definition videos, operating systems, and complex scientific databases into DNA [15].

Modern encoding techniques employ fountain codes, trellis codes, and specialized modulation schemes. Fountain codes, in particular, are highly resilient to data dropouts; they generate a limitless stream of data droplets from the source file, allowing the original data to be perfectly reconstructed even if a large subset of the synthesized DNA strands is degraded, mutated, or lost during bacterial extraction. In addition, advanced error-correction codes – such as Reed–Solomon codes or algorithms calculating Levenshtein distances – are used to detect and mathematically reverse biological mutations [16].

The decoding process has also improved significantly due to rapid advances in sequencing technologies and computational biology. High-throughput sequencing platforms generate millions of overlapping reads simultaneously, which are then aligned and analyzed. Modern decoding software utilizes consensus call algorithms that cross-reference multiple reads of the same sequence to statistically eliminate random single-nucleotide polymorphisms or sequencing artifacts. This computational back-end acts as a digital filter, guaranteeing that biological mutations inherent in living organisms do not translate into digital file corruption.

ROLE OF PLASMIDS AND CONJUGATION

Plasmids play a central role in bacterial information storage because they are relatively easy to isolate, manipulate *in vitro*, and transfer between different bacterial hosts. Scientists can engineer plasmids with customized multiple cloning sites (MCS) to carry specific DNA sequences representing massive blocks of encoded digital information. These “data plasmids” can be designed as shuttle vectors, allowing them to be propagated across different strains or species of bacteria depending on the environmental requirements of the storage facility.

Additionally, plasmids can be engineered with selective markers such as antibiotic resistance genes or essential metabolic complementation pathways. This ensures that only the bacteria carrying the digital data survive in a controlled cultivation environment. By applying continuous selective pressure (e.g., maintaining specific antibiotics in the culture medium), researchers prevent data loss via natural plasmid curing – a process where bacteria spontaneously expel plasmids that impose an unnecessary metabolic burden.

Bacterial conjugation provides a natural, highly efficient mechanism for transferring plasmids between cells. During conjugation, genetic material moves from a donor bacterium to a recipient bacterium through direct physical contact, facilitated by a specialized protein structure called a conjugation pilus. This phenomenon has inspired the concept of biological communication networks, or “living networks”, where bacteria act as microscopic carriers and transceivers of digital information.

In such a scenario, data could be propagated through a dense bacterial biofilm or liquid culture without requiring any electronic routing infrastructure, utilizing the natural metabolic energy of the cells to drive data distribution. Although current biological transfer rates are much slower than those achieved by modern electronic communication systems (such as Wi-Fi or fiber-optic networks), bacterial conjugation demonstrates the fundamental possibility of transmitting complex digital information securely through living organisms. Continued research into high-frequency recombination (Hfr) strains and optimized pheromone-inducible plasmid systems may significantly improve the speed and efficiency of such biological data transmission networks in the future, potentially enabling autonomous data routing within living ecosystems [17].

RECENT ADVANCES IN THE FIELD (2024–2026)

The field of DNA data storage has progressed rapidly over the last decade, transitioning from a theoretical novelty to a viable technological pipeline. While early research primarily established proof-of-concept for DNA storage, the literature from 2024 onward underscores a rapid evolution toward practical, scalable, and dynamic biological storage networks.

Improvements in DNA synthesis technologies – particularly the shift toward enzymatic DNA synthesis using Terminal Deoxynucleotidyl Transferase (TdT) – have significantly reduced monetary costs and increased production capacity compared to traditional phosphoramidite chemistry. Enzymatic methods bypass the need for hazardous organic solvents, function efficiently in aqueous environments, and enable the rapid synthesis of longer, more accurate oligonucleotides. At the same time, next-generation sequencing (NGS) platforms and ultra-portable nanopore sequencers have enhanced the speed, portability, and accuracy of information retrieval, allowing for rapid, field-ready decoding without reliance on massive laboratory infrastructure.

A major recent trend is the integration of CRISPR technologies for dynamic data manipulation and retrieval. Using CRISPR-Cas9 or CRISPR-Cas12 systems, scientists can now introduce information-bearing sequences into highly stable genomic loci with remarkable precision, avoiding the instability risks associated with free-floating plasmids [18]. Furthermore, cutting-edge innovations – such as the recent deployment of the SEEKER system, which utilizes CRISPR-Cas12a to rapidly perform quantitative keyword searches via target-specific fluorescence – have fundamentally addressed the historical challenge of performing random-access queries within complex DNA pools.

Synthetic biology has further expanded these possibilities by enabling the development of programmable genetic circuits and CRISPR-based molecular tickers. These living tickers can record

real-time biological events, metabolic fluctuations, and environmental signals (such as toxin exposure or radiation) directly into the host bacterium's genome, functioning as an autonomous cellular black box. Additionally, the advent of "epi-bit" technology in late 2024 has introduced novel epigenetic encoding methods, radically expanding in vivo data density by utilizing methylation patterns to store temporary digital states without altering the foundational base sequence.

To manage this immense scale, artificial intelligence and machine learning are increasingly being integrated into the workflow. AI models are used to optimize encoding strategies, predict DNA secondary structure formations that might hinder synthesis, and improve retrieval efficiency. Pivotal 2025 breakthroughs, such as the AI-driven "DNAformer" method, exemplify this shift by deploying deep learning models to intelligently reconstruct noisy sequencing reads, accelerating DNA sequence retrieval speeds by several magnitudes while maintaining strict bit-perfect accuracy. This convergence of deep learning and synthetic biology has drastically accelerated the development of high-density in vivo storage architectures.

ADVANTAGES OF BACTERIAL DATA STORAGE

Bacterial data storage offers several significant advantages over conventional silicon and magnetic infrastructure. The most obvious is exceptional storage density. Because DNA packages information at the molecular level, it can store enormous amounts of information in a remarkably small physical volume. A single milliliter of bacterial culture can contain billions of individual cells, each acting as an independent storage drive. This allows for the consolidation of entire warehouse-sized data centers into small, easily maintainable biological repositories, directly mitigating the immense real estate and material demands of modern digital storage infrastructure [19].

Another critical advantage is long-term durability. Under suitable conditions, such as dehydration or spore formation, bacterial hosts can shield DNA from environmental stressors, allowing it to remain stable for extended periods. When data is stored inside bacterial spores (such as those of *Bacillus subtilis*), the cellular architecture condenses and forms a highly robust, multi-layered protective shell. This makes the engineered bacteria hyper-resistant to extreme heat, ionizing radiation, and severe desiccation, offering a highly robust solution for ultra-long-term archival preservation.

This biological resilience offers a profound shift for sustainable data archiving, effectively decoupling data longevity from active energy consumption. Unlike many electronic devices that require continuous power, sophisticated climate control, and regular hardware replacements to prevent catastrophic data loss, biological storage systems preserve information with an absolute zero energy footprint during the cold-storage phase.

Bacteria also possess autonomous self-replicating capabilities. As cells divide, they automatically generate additional identical copies of the stored information using native, highly energy-efficient metabolic processes. This massively reduces the need for conventional electronic backup systems and manual server duplication. Furthermore, biological storage technologies have the potential to significantly reduce global dependence on rare-earth metals, toxic flame retardants, and short-lived, non-recyclable electronic components. By transitioning digital archives to organic, biodegradable mediums, bacterial storage contributes to a much more sustainable, eco-friendly approach to global information management that perfectly aligns with modern green computing initiatives [20].

CHALLENGES AND LIMITATIONS

Despite its unprecedented density, sustainability, and considerable promise, rigorous critical evaluation reveals severe biophysical and economic limitations that currently prevent the immediate commercial adoption of bacterial data storage.

DNA synthesis and sequencing remain relatively expensive compared with conventional storage technologies like flash memory or magnetic tape. Although the cost of genome sequencing has dropped

at a rate faster than Moore's Law, the cost of custom de novo DNA synthesis remains a primary economic bottleneck. Writing gigabytes or terabytes of arbitrary digital data into precise nucleotide sequences requires considerable financial investment, currently limiting large-scale implementation to highly specialized proof-of-concept projects rather than consumer-level applications.

More fundamentally, retrieval latency presents a massive structural barrier. Reading information from DNA is considerably slower than accessing data from modern electronic devices, which operate on millisecond timescales. The data retrieval pipeline for bacterial storage requires cell lysis, chemical DNA extraction, sequencing library preparation, and complex computational sequence alignment – a process that can take anywhere from several hours to days. Consequently, bacterial storage is currently entirely unsuitable for hot storage, random-access applications, or everyday computing tasks. It is, for the foreseeable future, uniquely positioned only for cold, long-term archival applications where data is rarely accessed but must be preserved securely for decades.

From a biological standpoint, the volatility of living systems poses a critical risk to data integrity. Genetic mutations represent an inherent biological vulnerability to stored digital information. As bacteria replicate over hundreds of generations, natural replication errors, environmental radiation, and evolutionary pressures can introduce point mutations, insertions, or deletions into the data-bearing sequences. If a mutation occurs within a critical data block, it can irreversibly corrupt the file. While researchers employ advanced redundancy and error-correction systems to minimize these risks, the core assumption that endogenous repair mechanisms will perpetually protect synthetic sequences must be rigorously challenged. Evolutionary pressure inherently works to eliminate non-functional artificial DNA that drains metabolic energy, representing a significant gap in our understanding of ultra-long-term genomic stability. Furthermore, complex regulatory, ethical, and biosafety considerations regarding the containment and environmental release of data-bearing genetically modified organisms (GMOs) must be comprehensively addressed before widespread industrial adoption.

APPLICATIONS

Potential applications of bacterial data storage are diverse, far-reaching, and highly disruptive to current technological paradigms. By transcending the physical limitations of silicon, this technology holds the potential to permanently redefine human knowledge preservation. Governments and national institutions could use biological systems for the secure archival preservation of critical historical records, legal constitutions, land registries, and cultural masterpieces. Storing these assets in biologically resilient spores ensures they can survive severe global catastrophes, electromagnetic pulse (EMP) attacks, or catastrophic grid failures.

Scientific organizations, such as CERN or major astronomical observatories, may employ DNA storage to maintain massive, multi-terabyte research datasets that are currently exorbitantly expensive to host on active cloud servers. Healthcare systems could preserve comprehensive patient medical records, life-long genomic histories, and complex epidemiological data in highly compact biological formats within non-pathogenic commensal bacteria safely stored in hospital bio-vaults.

Space exploration represents another exceptionally promising application. Because DNA offers unmatched storage density and durability under extreme conditions, it can provide an ultra-lightweight, highly efficient means of preserving massive amounts of technical data, scientific maps, and structural blueprints during long-duration interstellar missions. In rocketry, where every gram of payload is critical and heavily dictates fuel costs, swapping heavy silicon drives for a milligram of bacterial spores could revolutionize deep-space data transmission.

Furthermore, cultural heritage preservation, localized environmental monitoring via smart bacterial sensors, and molecular computing – where data storage and complex logical processing happen simultaneously within the same biological matrix – are additional areas where bacterial storage technologies are poised to prove invaluable.

FUTURE PERSPECTIVES

The future of bacterial data storage appears highly promising, driven by the rapid interdisciplinary convergence of computer science, molecular biology, and advanced chemistry. Continued advances in synthetic biology, nanotechnology, artificial intelligence, and automated genome engineering are expected to systematically improve performance, increase write/read throughput and drastically reduce the commercial costs of enzymatic synthesis.

A crucial next step in the evolution of this field is transitioning from passive biological archives to active “living computers”. Researchers envision future hybrid storage systems that seamlessly combine the high processing speed of electronic silicon computing with the unmatched density and longevity of biological storage. By utilizing automated microfluidic interfaces and lab-on-a-chip technologies, engineers hope to bridge the digital and biological domains, creating server racks where robotic fluid handlers automatically manage liquid bacterial cultures.

Future developments may also enable living cells to function as fully dynamic, rewritable memory systems similar to random access memory (RAM). By engineering reversible recombinase switches and synthetic epigenetic markers, scientists are pushing to create bacteria that can not only store static historical data but actively record, modify, and erase information in response to real-time environmental and biological events over extended temporal scales. As sequencing technologies become faster, more automated, and more affordable, these intelligent bacterial storage systems will likely transition from experimental laboratory concepts to highly practical, commercially viable archival solutions that fundamentally change the architecture of data centers worldwide.

CONCLUSION

Bacterial data storage represents a fascinating and immensely powerful convergence of microbiology, genetics, and information technology. By utilizing DNA as a highly stable, ultra-dense storage medium, researchers have successfully demonstrated the theoretical and practical possibility of preserving complex digital information within living organisms. The transition from binary electronic data to quaternary biological sequences opens new frontiers in data density, sustainability, and longevity that silicon-based technologies simply cannot match.

Although significant technical, economic, and regulatory challenges remain – particularly regarding high de novo synthesis costs, substantial read latency, and the ongoing battle against natural biological mutation rates – the remarkable storage density, long-term environmental durability, and autonomous self-replicating nature of bacterial systems make them incredibly attractive candidates for the future of archival storage. Continued, intensive interdisciplinary research will be absolutely critical in optimizing these biological platforms, mitigating current biophysical limitations, and ultimately transforming this innovative conceptual breakthrough into a practical, globally scalable technology capable of safeguarding the entirety of human knowledge for generations to come.

REFERENCES

1. Buko T, Tuczko N, Ishikawa T. DNA data storage. *BioTech (Basel)*. 2023;12(2):44. doi: [10.3390/biotech12020044](https://doi.org/10.3390/biotech12020044).
2. Wang S, Mao X, Wang F, Zuo X, Fan C. Data storage using DNA. *Adv Mater*. 2024;36(6):e2307499. doi: [10.1002/adma.202307499](https://doi.org/10.1002/adma.202307499). PMID:37800877.
3. Jo S, Shin H, Joe SY, Baek D, Park C, Chun H. Recent progress in DNA data storage based on high-throughput DNA synthesis. *Biomed Eng Lett*. 2024;14(5):993–1009.
4. Şatır E. A DNA data storage method using spatial encoding based lossless compression. *Entropy (Basel)*. 2024;26(12):1116.
5. Bar-Lev D, Sabary O, Yaakobi E. The zettabyte era is in our DNA. *Nat Comput Sci*. 2024;4(11):813–817.
6. Organick L, Ang SD, Chen YJ, Lopez R, Yekhanin S, Makarychev K, et al. Random access in large-scale DNA data storage. *Nat Biotechnol*. 2018;36(3):242–248.
7. Ceze L, Nivala J, Strauss K. Molecular digital data storage using DNA. *Nat Rev Genet*. 2019;20(8):456–466.

8. Goldman N, Bertone P, Chen S, Dessimoz C, LeProust EM, Sipos B, et al. Towards practical, high-capacity, low-maintenance information storage in synthesized DNA. *Nature*. 2013;494(7435):77–80.
9. Church GM, Gao Y, Kosuri S. Next-generation digital information storage in DNA. *Science*. 2012;337(6102):1628.
10. Shipman SL, Nivala J, Macklis JD, Church GM. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*. 2017;547(7663):345–349.
11. Erlich Y, Zielinski D. DNA Fountain enables a robust and efficient storage architecture. *Science*. 2017;355(6328):950–954.
12. Ping Z, Chen S, Zhou G, Huang X, Zhu SJ, Zhang H, et al. Towards practical and robust DNA-based data archiving using the yin-yang codec system. *Nat Comput Sci*. 2022;2(4):234–242. doi: [10.1038/s43588-022-00231-2](https://doi.org/10.1038/s43588-022-00231-2). PMID:38177542. PMCID:PMC10766522.
13. Heckel R, Shomorony I, Ramchandran K, David. Fundamental limits of DNA storage systems. arXiv. 2017. Available from: <https://arxiv.org/abs/1705.04732>
14. Doricchi A, Platnich CM, Gimpel A, Horn F, Earle M, Lanzavecchia G, et al. Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS Nano*. 2022 Oct 18;16(11):17552–71. Available from: <https://pubs.acs.org/ancac3/article/16/11/17552/1352375/Emerging-Approaches-to-DNA-Data-Storage-Challenges>
15. Takahashi CN, Nguyen BH, Strauss K, Ceze L. Demonstration of end-to-end automation of DNA data storage. *Sci Rep*. 2019;9(1):4998.
16. Wang Y, Noor-A-Rahim M, Zhang J, Gunawan E, Guan YL, Poh CL. High capacity DNA data storage with variable-length oligonucleotides using repeat accumulate code and hybrid mapping. *J Biol Eng*. 2019;13(1):89.
17. Cho S, Shin J, Cho BK. Applications of CRISPR/Cas system to bacterial metabolic engineering. *Int J Mol Sci*. 2018;19(4):1089. doi: [10.3390/ijms19041089](https://doi.org/10.3390/ijms19041089). PMID:29621180. PMCID:PMC5979482.
18. O'Driscoll A, Sleator RD. Synthetic DNA: The next generation of big data storage. *Bioengineered*. 2013;4(3):123–125. doi: [10.4161/bioe.24296](https://doi.org/10.4161/bioe.24296). PMID:23514938. PMCID:PMC3669150.
19. Smith LN, Raghav NK. Recombinant plasmid vectors and their industrial stability footprints. *J Pharm Appl Sci*. 2025;14(1):89–101.
20. Zhirnov V, Zadegan RM, Sandhu GS, Church GM, Hughes WL. Nucleic acid memory. *Nat Mater*. 2016;15(4):366–370.