

DNA Information Storage and Cryptography System

Zeping Zhang^{1, *}, Zhihao Zhang¹

¹ School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China

* Corresponding author: Zeping Zhang (Email: 120212227312@ncepu.edu.cn)

Abstract: With the development of information technology, the global data volume is growing exponentially. In order to alleviate the contradiction between massive data and traditional storage technology, people begin to seek for a new generation of storage media. As a carrier of genetic information, DNA has the characteristics of high information density, long storage life and low maintenance cost, which can effectively overcome the deficiency of traditional storage media. With the development of DNA synthesis and DNA sequencing technology, DNA data storage technology has attracted more and more attention, and a series of major breakthroughs have been made. In this paper, with the workflow of DNA data storage as the main line, expounds the basic theory of DNA data storage and related technology, mainly introduced the research progress of DNA storage method and strategy, briefly summarizes the latest research results of DNA data cryptography, and finally discussed the major challenges that DNA data storage technology is facing, especially, DNA synthesis efficiency, DNA sequencing time cost and DNA data cryptography will be an important research direction of DNA storage technology in the future. It is believed that with the deepening of the data storage research on DNA, DNA data storage will become the most potential new storage method in the future, and can be a practical application in the future.

Keywords: DNA information storage; DNA synthesis; DNA preservation; DNA sequencing; DNA cryptography.

1. Introduction

In the 21st century, with the rapid development of information technologies such as 5G, the Internet of Things, and artificial intelligence, the amount of information is growing exponentially, traditional storage methods are gradually unable to meet the need[1,2]. According to the Internet Data Center (IDC), in 2025, the global data volume will reach 175 ZB (Ze bytes), with a five-year compound annual growth rate of 31.8%, which will far exceed the storage capacity of any currently available storage methods[3–5]. In response to this growth, the costs of maintaining and transmitting data, limited longevity and significant data loss, a new method of information storage is urgently needed[6,7]. DNA is an engineered chemical that can be used to build novel storage systems due to its predictable Watson Crick base pairing principles[8] and extremely high data storage density.

As a material that carries genetic information, DNA is theoretically the most suitable medium for molecular level digital information storage[9–12]. Compared with traditional storage media, DNA has high information density, long storage life and low maintenance cost, which has great development potential[13–15]. Although many strategies using organic molecules for digital information storage have been proposed[16–20], the use of DNA molecules for storing digital information remains the most widely accepted strategy due to the cost and throughput advantages of current DNA sequencing technologies[21–26].

In recent years, significant progress has been made in using DNA as a digital information storage medium[27–34]. The existing DNA storage strategy is mainly divided into the following steps. First, the information to be stored is encoded into DNA sequence by using DNA synthesis technology, DNA is stored in vivo or in vitro conditions, and specific DNA sequence can be randomly accessed according to user's needs. Second, DNA sequence information is read through DNA sequencing technology. Finally, the sequencing results of

DNA sequence are decoded into stored information. The rapid development of DNA data storage technology has benefited from the tremendous advances in biotechnology over the past few decades[35–38]. These biotechnology include enzymatic DNA synthesis, polymerase chain reaction for DNA amplification and DNA sequencing technology[39–43].

In this review, we take the process of DNA storage as the main line, and systematically explain the following aspects: (1) the research progress of DNA coding technology; (2) the development of DNA synthesis technology; (3) the methods and strategies of DNA preservation; (4) the latest progress of DNA sequencing technology; (5) DNA cryptography and DNA data encryption technology. Finally, the main challenges and development trends of DNA data storage at this stage are discussed, hoping that the development of DNA storage technology can be promoted through this review.

2. The Research Studies on DNA Storage

DNA storage, a technology that can store information in DNA, was first proposed in 1959 by the American physicist Feynman. In DNA storage system, information is first encoded as binary or quaternary(base arrangement) information, which can be stored in DNA sequence[44–48] or DNA origami[49,50] through DNA synthesis technology. After a long period of DNA preservation, DNA sequence will be measured by DNA sequencing technology[51], the original information can be obtained by decoding DNA sequence using coding system.

2.1. Coding Technology of DNA Storage

The simplest coding method is to use A, T, C, G corresponding to 00,01,10,11. In this way, DNA storage logic density[53] can theoretically reach 2bits/nt (bit per nucleotide), but the mapping method will have problems such as single base repeat, CG content imbalance, uncontrollable DNA secondary structure, DNA stability[30] and so on.

In order to solve these problems, George Church mapped A and C to 0, and G and T to 1, solving the imbalance of CG content and other issues. However, the cost was to reduce the DNA storage density, which theoretically can only reach 1 bit/nt[13]. In order to further improve the storage density, Goldman first conducts ternary Huffman coding[54] for the binary information, and then conducts rotation coding for the ternary information. In this way, problems such as single base repeat and CG content imbalance are avoided, and the theoretical storage density reaches 1.58bits/nt[55]. Due to the molecular bias[56], Grass uses RS code[57] to add error correction mechanism[58] to DNA data storage[59], and improves the theoretical storage density to 1.78bits/nt on the premise of improving the accuracy[60]. Zhi Ping et al. designed the Yin Yang code inspired by Yin Yang and five elements theory. They used two coding systems to code simultaneously, cleverly avoiding issues such as single base repeat, CG content imbalance, and uncontrollable DNA secondary structures in a dual coding system, They also increased the theoretical logic density to 1.95 bits/nt[61].

Leon Anavy reduced the average number of DNA synthesis cycles and greatly improved the actual storage density through probability recognition through the use of composite DNA letters. Because it is impossible to only synthesize one strand of DNA during DNA synthesis, this DNA storage encoding system actually defines new letters by identifying the nucleotide ratio of the same DNA, greatly improving the actual storage density in way that defining new composite DNA letters[62]. Erlich designed a set of DNA fountain codes by using for reference of fountain codes[63] from coding theory, which store complete information by dividing it into several droplets, much like a fountain. Increase the theoretical storage density to 1.98 bits/nt, which is extremely close to the theoretical limit of 2 bits/nt, but the drawback is that once some data is missing, the original information cannot be restored and some special binary sequences cannot be encoded successfully[64].

2.2. DNA Synthesis Technology

Any type of file such as text and pictures can be represented as a bit sequence. DNA storage is essentially the use of DNA sequences. While the restriction on the length of the coding DNA sequence comes from the chemical synthesis of DNA, which is prone to generate errors in the DNA sequence when synthesized over a few hundred nucleotides.

Artificial DNA synthesis is a technique for DNA synthesis artificially designed from an arbitrary sequence without relying on a DNA template. DNA chemical synthesis technology started in the late 1940s. In 1953, Watson and Crick revealed the double helical structure of DNA, and this breakthrough work not only revealed the structure of DNA, but also opened the door for subsequent synthetic biology studies[65]. In 1955, the Todd Laboratory of Cambridge University successfully synthesized TpT with a phosphodiester bond structure, which won the Nobel Prize in 1957[66]. In 1965, Khorana and his team synthesized nucleotides and gene codons through organic chemical synthesis methods, which won the Nobel Prize in 1968. In 1981, Marvin Caruthers proposed the method for the chemical synthesis of DNA using phosphor amidite intermediates[67]. This is a four step cycle reaction that involves adding the desired nucleotide to a growing oligonucleotide fixed on a solid support material[68], using solid support is for the achievement of extensive parallel

synthesis, and automated of chemical processes[69,70].

Although there are many advantages of chemical synthetic, it is noteworthy that the use of this method requires toxic chemical reagents in the synthesis process. In order to reduce the use of chemical reagents and organic solvents with potential negative environmental impact, researchers have tried to develop synthesis methods that do not rely on toxic chemical reagents. Enzymatic DNA synthesis technologies and electrochemical synthesis technologies have been rapidly developed, especially for template-independent enzyme oligonucleotide synthesis (TiEOS), which uses terminal deoxynucleotidyl transferase (TdT) for DNA synthesis[71]. In 2021, Eojin Yoo et al found that methods relying on T4 rnl ligase or TdT enzymes could be used to add bases specifically to growing oligonucleotides in an aqueous environment, thus eliminating the requirement for organic solvents[72]. Liu Hong's team from Southeast University improved traditional chemical synthesis methods of phosphor amides through electrochemical deprotection technology, and sequenced the DNA molecules on the electrode surface based on the charge oscillation phenomenon, and invented a DNA storage system based on electrochemical method[35].

2.3. Methods of DNA Preservation

Compared to other storage media, a significant advantage of DNA data storage is the ability to improve data retention time through DNA preservation technology. However, naturally unprotected DNA is very fragile, vulnerable to hydrolyze or be oxidized, and has a characteristic half-life period[73,74]. Half-life period of DNA is closely related to the storage temperature and the length of the DNA chains, low temperatures and waterproof environments can significantly improve their stability. For example, DNA solution can usually be stable at room temperature for 3 to 6 months, while it can be stored for about 1 year at 4 degrees centigrade and 2 years under freezing conditions at -20 degrees centigrade[75]. In order to realize long-term data storage more effectively, researchers have developed a variety of methods and strategies for DNA preservation.

In order to prevent sample degradation during transportation, storage, and processing, because of the difficulty of long-term preservation of DNA solution, people usually freeze dry and dehydrate DNA molecules for preservation, this is not only suitable for long-term stable preservation of samples, but also for fast and complete sample recovery. Sharon Newman et al. arranged dehydrated DNA spots densely on glass plates, then the glass plates were placed on digital microfluidic equipment to retrieve data, and successfully realized a method of DNA data storage based on digital microfluidic technology[76].

DNA molecules can be preserved in skeletal debris or sediment for hundreds of thousands of years because the dense outer layer of skeletal debris or sediment separates DNA from the water and reactive oxygen species in the environment[77,78]. Inspired by this, the scientific researchers have conducted new research. Chunhai Fan et al. used the nucleic acid frame structure as the template and the electrostatic adsorption as the driving force, and successfully prepared the calcium phosphate nanocrystals with highly controllable geometry[79], DNA stability is greatly enhanced due to the isolated and protective effect of the outer calcium phosphate. Daniela Paunescu et al. encapsulated DNA in silica particles, mimicking fossils to protect DNA from corrosive environments. DNA is immobilized on the surface

of cationic charged silica particles, on which TEOS deposits a dense silica layer. TMAPS is used as a co interacting material, realizing the compatibility between the sol gel process and DNA[80]. Puddu has developed a core-shell protective structure. The introduction of magnetic cores enables the protection carrier to aggregate under a magnetic field, promoting swift information recovery. By combining multiple cations with DNA molecules, a layer by layer encapsulation method is achieved[81]. Kohll developed a DNA encapsulation method with high DNA loading and simple sample processing properties by simulating fossil components. Compared to other DNA storage system, earth-alkaline salt storage system is uncomplicated, swift, easily automated, and maintain remarkable DNA stabilization at very high DNA load, even at 50% relative humidity[82].

2.4. DNA Sequencing Technology

For reading out large amount of information stored in DNA[83], DNA sequencing technology is needed. DNA sequencing was first implemented by Frederick Sanger, using DNA polymerase to extend primers bound to the undetermined sequence template until a strand termination nucleotide was introduced[84]. For improving the throughput of sequencing (the total number of sequences that can be measured in each experiment), humans began to develop next-generation sequencing technology[85–91]. The first two generations of sequencing were both techniques of synthesis while sequencing, which refers to the technique of sequencing by measuring the nucleotides added to each new strand during the DNA replication process. To achieve single molecule sequencing and obtain ultra long sequencing read length, nanopore sequencing[92] and single molecule real-time(SMRT) sequencing[93] have been invented. SMRT can be used for phased diploid genome assembly[94] and direct detection of DNA methylation[95]. Nanopore sequencing is based on nanopore electrical signal sequencing technology[96], which utilizes a nanopore with covalent molecular junctions to fix the nanopore protein onto a resistive membrane, and then pull the nucleic acid through the nanopore by protein. When a single base passes through a nanoscale channel, it will cause changes in the electrical properties of the channel. In theory, the differences in the chemical properties of four different bases (A, C, G, T) can lead to different changes in electrical parameters when they pass through nanopore. Detecting these changes can obtain the corresponding types of bases, thereby achieving sequencing[92]. Nanopore can be also used for detection of microRNA, protein, small biomarkers[97] and direct observation of DNA knots[98]. In the field of DNA storage, nanopore can be used for DNA data storage readout[99].

3. The Research Studies on DNA Cryptography

The vast amount of information contained in DNA can also be used for encryption[100–103] and true random number generation[104]. In 1999, Zapp developed DNA based dual steganography technology in DNA microdots for sending secret messages. The information encoded by DNA is first disguised in the vast and complex genomic DNA of humans, and then further hidden and limited to DNA microdots[105].

In 2016, Clemens Mayer was inspired by epigenetic regulation of dynamic biological information discovered how binary data controls information when encoded in synthetic

DNA strands. Reactions of cytosines and their natural derivatives demonstrate how to store multilayer information in a single DNA template, which hides multiple information in the same DNA template, and demonstrate that controlled redox reactions allow the mutual transformation of information layers encoded in DNA. Overall, such storage of multiple pieces of information in a single synthetic individual DNA library demonstrates the latent capacity of chemical reactions in processing digital information of biopolymers[20].

Jangwon Kim found that the chemical stability of DNA posed difficulties in completely deleting the information encoded in DNA sequences. Therefore, he encoded the information as a mixture of oligonucleotides encoded by a mixture of true and false information, which could quickly and permanently erase the information. The true information is distinguished by hybridization with oligonucleotides labeled as "real", and can only read the real information sequence; Even brief exposure to high temperatures can effectively randomize binding with real markers. 8 independent point bitmap images were shown to be able to be stably stored at 25 ° C for 65 days for reading, with an average correct information recovery rate of over 99%. It is inferred that the half-life at 25 ° C exceeds 15 years, but heating to 95 ° C for 5 minutes will permanently delete the message[34].

In 2019, Chunhai Fan's team created a unique information security method using biomolecular cryptography for data encryption through specific biomolecular interactions. However, the development of encryption protocols based on biomolecular reactions to ensure the confidentiality, integrity, and availability of information remains a significant challenge. In this study, they developed DNA Origami Cryptography (DOC), which utilizes M13 virus scaffolds to fold into nanoscale graphics for safety communication, creating keys exceeding 700 bits. The inherent nanoscale addressing ability of this DNA origami also enables protein-binding based steganography, enhancing the protection of information confidentiality and integrity within DOC. By creating unique connections between multiple DNA, information transmission can be ensured. Origami carries a portion of information. The versatility of DOC is further reflected in the transmission of diverse data formats, in order to meet the enormous potential of the rapidly growing encryption needs[106].

Robert N. Grass found a strategy to combine these two technologies by reading the human genome and storing digital data in synthesized DNA, in order to achieve valuable information in synthesized DNA. He discovered that genetic short tandem repeats (STR) sequences embody entropy keys sufficient to realize strong encryption. Using this method, an 80 bits strong key was generated from human DNA through experiments, and the information stored in synthetic DNA at 17KB was encrypted using such keys. Finally, the information is perfectly decrypted[107].

DNA offers a range of advanced tools that can be utilized in developing cryptographic methods as well as biosensors. However, traditional approaches to regulating DNA primarily focus on controlling enthalpy. This approach often leads to unpredictable responses to stimuli and less accurate outcomes due to significant energy fluctuations. To address these limitations effectively, Lin Lin Zheng devised a novel strategy based on both enthalpy and entropy regulations that respond specifically to changes in pH levels. These innovative DNA motifs have been successfully employed in glucose sensing

systems as well as crypto-steganography applications. Their successful implementation highlights their immense potential in both bio-sensing technologies and secure data encryption[103].

The security of modern cybersecurity systems, which rely on public-key cryptosystems like Rivest-Shamir-Adleman, can be compromised when solutions to prime factorization are discovered. Yanan Zhang has developed DNA origami frameworks (DOFs) to guide the localized assembly of double-crossover (DX) tiles for solving prime factorization. By utilizing a model comprising computing, decision-making, and reporting motifs, this DOF-based demonstration successfully achieves the factorization of semiprimes 6 and 15[108].

4. Conclusion

DNA storage technology is an epoch-making storage technology that focuses on the future. It uses artificially synthesized deoxyribonucleic acid (DNA) as a storage medium, which has the advantages of high efficiency, large storage capacity, long storage time, easy access, and maintenance free. The attraction of DNA for information storage lies in the extremely high information density generated by molecular scale information storage. However, there are currently some bottlenecks in DNA storage: high costs for DNA synthesis and sequencing, and slow DNA sequencing speed. If these problems can be solved, DNA storage technology and DNA Cryptography will take a leap forward.

References

- [1] Rydning D R J G J, Reinsel J, Gantz J. The Digitization of the World from Edge to Core[J]. Framingham: International Data Corporation, 2018, 16: 1-28.
- [2] Zhirmov, V.; Zadegan, R.M.; Sandhu, G.S.; Church, G.M.; Hughes, W.L. Nucleic Acid Memory. *Nature Materials* 2016, 15, 366–370.
- [3] D. Carmean; L. Ceze; G. Seelig; K. Stewart; K. Strauss; M. Willsey DNA Data Storage and Hybrid Molecular–Electronic Computing. *Proceedings of the IEEE* 2019, 107, 63–72.
- [4] K. Goda; M. Kitsuregawa The History of Storage Systems. *Proceedings of the IEEE* 2012, 100, 1433–1440.
- [5] Bhat, W.A. Bridging Data-Capacity Gap in Big Data Storage. *Future Generation Computer Systems* 2018, 87, 538–548.
- [6] Williams, E.D.; Ayres, R.U.; Heller, M. The 1.7 Kilogram Microchip: Energy and Material Use in the Production of Semiconductor Devices. *Environ. Sci. Technol.* 2002, 36, 5504–5510.
- [7] Andrae, A.S.G.; Edler, T. On Global Electricity Usage of Communication Technology: Trends to 2030. *Challenges* 2015, 6, 117–157.
- [8] Neidle, S.; Sanderson, M. Chapter 2 - The Building Blocks of DNA and RNA. In *Principles of Nucleic Acid Structure (Second Edition)*; Neidle, S., Sanderson, M., Eds.; Academic Press: New York, 2022; pp. 29–51.
- [9] Takahashi, C.N.; Nguyen, B.H.; Strauss, K.; Ceze, L. Demonstration of End-to-End Automation of DNA Data Storage. *Scientific Reports* 2019, 9, 4998.
- [10] Organick, L.; Ang, S.D.; Chen, Y.-J.; Lopez, R.; Yekhanin, S.; Makarychev, K.; Racz, M.Z.; Kamath, G.; Gopalan, P.; Nguyen, B.; et al. Random Access in Large-Scale DNA Data Storage. *Nature Biotechnology* 2018, 36, 242–248.
- [11] Ceze, L.; Nivala, J.; Strauss, K. Molecular Digital Data Storage Using DNA. *Nature Reviews Genetics* 2019, 20, 456–466.
- [12] Fei, Z.; Gupta, N.; Li, M.; Xiao, P.; Hu, X. Toward Highly Effective Loading of DNA in Hydrogels for High-Density and Long-Term Information Storage. *Science Advances* 9, eadg9933.
- [13] Church, G.M.; Gao, Y.; Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science* 2012, 337, 1628–1628.
- [14] Allentoft, M.E.; Collins, M.; Harker, D.; Haile, J.; Oskam, C.L.; Hale, M.L.; Campos, P.F.; Samaniego, J.A.; Gilbert, M.T.P.; Willerslev, E.; et al. The Half-Life of DNA in Bone: Measuring Decay Kinetics in 158 Dated Fossils. *Proceedings: Biological Sciences* 2012, 279, 4724–4733.
- [15] Dabney, J.; Knapp, M.; Glocke, I.; Gansauge, M.-T.; Weihmann, A.; Nickel, B.; Valdiosera, C.; García, N.; Pääbo, S.; Arsuaga, J.-L.; et al. Complete Mitochondrial Genome Sequence of a Middle Pleistocene Cave Bear Reconstructed from Ultrashort DNA Fragments. *Proceedings of the National Academy of Sciences* 2013, 110, 15758–15763.
- [16] Kennedy, E.; Arcadia, C.E.; Geiser, J.; Weber, P.M.; Rose, C.; Rubenstein, B.M.; Rosenstein, J.K. Encoding Information in Synthetic Metabolomes. *PLOS ONE* 2019, 14, e0217364.
- [17] Cafferty, B.J.; Ten, A.S.; Fink, M.J.; Morey, S.; Preston, D.J.; Mrksich, M.; Whitesides, G.M. Storage of Information Using Small Organic Molecules. *ACS Cent. Sci.* 2019, 5, 911–916.
- [18] Ng, C.C.A.; Tam, W.M.; Yin, H.; Wu, Q.; So, P.-K.; Wong, M.Y.-M.; Lau, F.C.M.; Yao, Z.-P. Data Storage Using Peptide Sequences. *Nature Communications* 2021, 12, 4242.
- [19] Lee, W.; Zhou, Z.; Chen, X.; Qin, N.; Jiang, J.; Liu, K.; Liu, M.; Tao, T.H.; Li, W. A Rewritable Optical Storage Medium of Silk Proteins Using Near-Field Nano-Optics. *Nature Nanotechnology* 2020, 15, 941–947.
- [20] Mayer, C.; McInroy, G.R.; Murat, P.; Van Delft, P.; Balasubramanian, S. An Epigenetics-Inspired DNA-Based Data Storage System. *Angewandte Chemie International Edition* 2016, 55, 11144–11148.
- [21] Larkin, J.; Henley, R.Y.; Jadhav, V.; Korlach, J.; Wanunu, M. Length-Independent DNA Packing into Nanopore Zero-Mode Waveguides for Low-Input DNA Sequencing. *Nature Nanotechnology* 2017, 12, 1169–1175.
- [22] Völler, J.-S. Enhancing DNA Sequencing. *Nature Catalysis* 2018, 1, 481–481.
- [23] Chen, Z.; Zhou, W.; Qiao, S.; Kang, L.; Duan, H.; Xie, X.S.; Huang, Y. Highly Accurate Fluorogenic DNA Sequencing with Information Theory–Based Error Correction. *Nature Biotechnology* 2017, 35, 1170–1178.
- [24] Nawy, T. Sequencing DNA, No Mistake. *Nature Methods* 2018, 15, 12–13.
- [25] Green, E.D.; Rubin, E.M.; Olson, M.V. The Future of DNA Sequencing. *Nature* 2017, 550, 179–181.
- [26] Maxam, A.M.; Gilbert, W. A New Method for Sequencing DNA. *Proceedings of the National Academy of Sciences* 1977, 74, 560–564.
- [27] Lim, C.K.; Nirantar, S.; Yew, W.S.; Poh, C.L. Novel Modalities in DNA Data Storage. *Trends in Biotechnology* 2021, 39, 990–1003.
- [28] Doricchi, A.; Platnich, C.M.; Gimpel, A.; Horn, F.; Earle, M.; Lanzavecchia, G.; Cortajarena, A.L.; Liz-Marzán, L.M.; Liu, N.; Heckel, R.; et al. Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS Nano* 2022, 16, 17552–17571.

- [29] Dong, Y.; Sun, F.; Ping, Z.; Ouyang, Q.; Qian, L. DNA Storage: Research Landscape and Future Prospects. *National Science Review* 2020, 7, 1092–1107.
- [30] Matange, K.; Tuck, J.M.; Keung, A.J. DNA Stability: A Central Design Consideration for DNA Data Storage Systems. *Nature Communications* 2021, 12, 1358.
- [31] Meiser, L.C.; Nguyen, B.H.; Chen, Y.-J.; Nivala, J.; Strauss, K.; Ceze, L.; Grass, R.N. Synthetic DNA Applications in Information Technology. *Nature Communications* 2022, 13, 352.
- [32] Koch, J.; Gantenbein, S.; Masania, K.; Stark, W.J.; Erlich, Y.; Grass, R.N. A DNA-of-Things Storage Architecture to Create Materials with Embedded Memory. *Nature Biotechnology* 2020, 38, 39–43.
- [33] Tomek, K.J.; Volkel, K.; Indermaur, E.W.; Tuck, J.M.; Keung, A.J. Promiscuous Molecules for Smarter File Operations in DNA-Based Data Storage. *Nature Communications* 2021, 12, 3518.
- [34] Kim, J.; Bae, J.H.; Baym, M.; Zhang, D.Y. Metastable Hybridization-Based DNA Information Storage to Allow Rapid and Permanent Erasure. *Nature Communications* 2020, 11, 5008.
- [35] Xu, C.; Ma, B.; Gao, Z.; Dong, X.; Zhao, C.; Liu, H. Electrochemical DNA Synthesis and Sequencing on a Single Electrode with Scalability for Integrated Data Storage. *Science Advances* 2020, 6, eabk0100.
- [36] Lee, H.H.; Kalthor, R.; Goela, N.; Bolot, J.; Church, G.M. Terminator-Free Template-Independent Enzymatic DNA Synthesis for Digital Information Storage. *Nature Communications* 2019, 10, 2383.
- [37] Lee, H.; Wiegand, D.J.; Griswold, K.; Punthambaker, S.; Chun, H.; Kohman, R.E.; Church, G.M. Photon-Directed Multiplexed Enzymatic DNA Synthesis for Molecular Digital Data Storage. *Nature Communications* 2020, 11, 5246.
- [38] Kubista, M.; Andrade, J.M.; Bengtsson, M.; Forootan, A.; Jonák, J.; Lind, K.; Sindelka, R.; Sjöback, R.; Sjögreen, B.; Strömbom, L.; et al. The Real-Time Polymerase Chain Reaction. *Molecular Aspects of Medicine* 2006, 27, 95–125.
- [39] Ralec, C.; Henry, E.; Lemor, M.; Killelea, T.; Henneke, G. Calcium-Driven DNA Synthesis by a High-Fidelity DNA Polymerase. *Nucleic Acids Research* 2017, 45, 12425–12440.
- [40] Kishi, J.Y.; Schaus, T.E.; Gopalkrishnan, N.; Xuan, F.; Yin, P. Programmable Autonomous Synthesis of Single-Stranded DNA. *Nature Chemistry* 2018, 10, 155–164.
- [41] Jiang, W.; Zhang, B.; Fan, C.; Wang, M.; Wang, J.; Deng, Q.; Liu, X.; Chen, J.; Zheng, J.; Liu, L.; et al. Mirror-Image Polymerase Chain Reaction. *Cell Discovery* 2017, 3, 17037.
- [42] Zhan, Y.; Zhang, J.; Yao, S.; Luo, G. High-Throughput Two-Dimensional Polymerase Chain Reaction Technology. *Anal. Chem.* 2020, 92, 674–682.
- [43] Heerema, S.J.; Dekker, C. Graphene Nanodevices for DNA Sequencing. *Nature Nanotechnology* 2016, 11, 127–136.
- [44] Sadremomtaz, A.; Glass, R.F.; Guerrero, J.E.; LaJeunesse, D.R.; Josephs, E.A.; Zedegan, R. Digital Data Storage on DNA Tape Using CRISPR Base Editors. *Nature Communications* 2023, 14, 6472.
- [45] Lin, K.N.; Volkel, K.; Tuck, J.M.; Keung, A.J. Dynamic and Scalable DNA-Based Information Storage. *Nature Communications* 2020, 11, 2981.
- [46] Song, L.; Geng, F.; Gong, Z.-Y.; Chen, X.; Tang, J.; Gong, C.; Zhou, L.; Xia, R.; Han, M.-Z.; Xu, J.-Y.; et al. Robust Data Storage in DNA by de Bruijn Graph-Based de Novo Strand Assembly. *Nature Communications* 2022, 13, 5361.
- [47] Heckel, R.; Mikutis, G.; Grass, R.N. A Characterization of the DNA Data Storage Channel. *Scientific Reports* 2019, 9, 9663.
- [48] Li, M.; Wu, J.; Dai, J.; Jiang, Q.; Qu, Q.; Huang, X.; Wang, Y. A Self-Contained and Self-Explanatory DNA Storage System. *Scientific Reports* 2021, 11, 18063.
- [49] Dey, S.; Fan, C.; Gothelf, K.V.; Li, J.; Lin, C.; Liu, L.; Liu, N.; Nijenhuis, M.A.D.; Saccà, B.; Simmel, F.C.; et al. DNA Origami. *Nature Reviews Methods Primers* 2021, 1, 13.
- [50] Dickinson, G.D.; Mortuza, G.M.; Clay, W.; Piantanida, L.; Green, C.M.; Watson, C.; Hayden, E.J.; Andersen, T.; Kuang, W.; Graunard, E.; et al. An Alternative Approach to Nucleic Acid Memory. *Nature Communications* 2021, 12, 2371.
- [51] Nguyen, B.H.; Takahashi, C.N.; Gupta, G.; Smith, J.A.; Rouse, R.; Berndt, P.; Yekhanin, S.; Ward, D.P.; Ang, S.D.; Garvan, P.; et al. Scaling DNA Data Storage with Nanoscale Electrode Wells. *Science Advances* 2020, 6, eabi6714.
- [52] Meiser, L.C.; Antkowiak, P.L.; Koch, J.; Chen, W.D.; Kohll, A.X.; Stark, W.J.; Heckel, R.; Grass, R.N. Reading and Writing Digital Data in DNA. *Nature Protocols* 2020, 15, 86–101.
- [53] Yan, Y.; Pinnamaneni, N.; Chalapati, S.; Crosbie, C.; Appuswamy, R. Scaling Logical Density of DNA Storage with Enzymatically-Ligated Composite Motifs. *Scientific Reports* 2023, 13, 15978.
- [54] D. A. Huffman A Method for the Construction of Minimum-Redundancy Codes. *Proceedings of the IRE* 1952, 40, 1098–1101.
- [55] Goldman, N.; Bertone, P.; Chen, S.; Dessimoz, C.; LeProust, E.M.; Sipos, B.; Birney, E. Towards Practical, High-Capacity, Low-Maintenance Information Storage in Synthesized DNA. *Nature* 2013, 494, 77–80.
- [56] Chen, Y.-J.; Takahashi, C.N.; Organick, L.; Bee, C.; Ang, S.D.; Weiss, P.; Peck, B.; Seelig, G.; Ceze, L.; Strauss, K. Quantifying Molecular Bias in DNA Data Storage. *Nature Communications* 2020, 11, 3264.
- [57] G. Solomon Self-Synchronizing Reed-Solomon Codes (Corresp.). *IEEE Transactions on Information Theory* 1968, 14, 608–609.
- [58] Welzel, M.; Schwarz, P.M.; Löchel, H.F.; Kabdullayeva, T.; Clemens, S.; Becker, A.; Freisleben, B.; Heider, D. DNA-Aeon Provides Flexible Arithmetic Coding for Constraint Adherence and Error Correction in DNA Storage. *Nature Communications* 2023, 14, 628.
- [59] Gimpel, A.L.; Stark, W.J.; Heckel, R.; Grass, R.N. A Digital Twin for DNA Data Storage Based on Comprehensive Quantification of Errors and Biases. *Nature Communications* 2023, 14, 6026.
- [60] Grass, R.N.; Heckel, R.; Puddu, M.; Paunescu, D.; Stark, W.J. Robust Chemical Preservation of Digital Information on DNA in Silica with Error-Correcting Codes. *Angewandte Chemie International Edition* 2015, 54, 2552–2555.
- [61] Ping, Z.; Chen, S.; Zhou, G.; Huang, X.; Zhu, S.J.; Zhang, H.; Lee, H.H.; Lan, Z.; Cui, J.; Chen, T.; et al. Towards Practical and Robust DNA-Based Data Archiving Using the Yin–Yang Codec System. *Nature Computational Science* 2022, 2, 234–242.
- [62] Anavy, L.; Vaknin, I.; Atar, O.; Amit, R.; Yakhini, Z. Data Storage in DNA with Fewer Synthesis Cycles Using Composite DNA Letters. *Nature Biotechnology* 2019, 37, 1229–1236.
- [63] MacKay, D.J.C. Fountain Codes. *IEE Proceedings - Communications* 2005, 152, 1062–1068(6).
- [64] Erlich, Y.; Zielinski, D. DNA Fountain Enables a Robust and Efficient Storage Architecture. *Science* 2017, 355, 950–954.

- [65] WATSON, J.D.; CRICK, F.H.C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* 1974, 248, 765–765.
- [66] Lin, Xi. Oligodeoxynucleotide Synthesis Using Protecting Groups and a Linker Cleavable under Non-Nucleophilic Conditions, Dissertation, Michigan Technological University, 2013.
- [67] Beaucage, S.L.; Caruthers, M.H. Deoxynucleoside Phosphoramidites—A New Class of Key Intermediates for Deoxypolynucleotide Synthesis. *Tetrahedron Letters* 1981, 22, 1859–1862.
- [68] Palluk, S.; Arlow, D.H.; de Rond, T.; Barthel, S.; Kang, J.S.; Bector, R.; Baghdassarian, H.M.; Truong, A.N.; Kim, P.W.; Singh, A.K.; et al. De Novo DNA Synthesis Using Polymerase-Nucleotide Conjugates. *Nature Biotechnology* 2018, 36, 645–650.
- [69] Kosuri, S.; Church, G.M. Large-Scale de Novo DNA Synthesis: Technologies and Applications. *Nature Methods* 2014, 11, 499–507.
- [70] LeProust, E.M.; Peck, B.J.; Spirin, K.; McCuen, H.B.; Moore, B.; Namsaraev, E.; Caruthers, M.H. Synthesis of High-Quality Libraries of Long (150mer) Oligonucleotides by a Novel Depurination Controlled Process. *Nucleic Acids Research* 2010, 38, 2522–2540.
- [71] Jensen, M.A.; Davis, R.W. Template-Independent Enzymatic Oligonucleotide Synthesis (TiEOS): Its History, Prospects, and Challenges. *Biochemistry* 2018, 57, 1821–1832.
- [72] Yoo, E.; Choe, D.; Shin, J.; Cho, S.; Cho, B.-K. Mini Review: Enzyme-Based DNA Synthesis and Selective Retrieval for Data Storage. *Computational and Structural Biotechnology Journal* 2021, 19, 2468–2476.
- [73] Baoutina, A.; Bhat, S.; Partis, L.; Emslie, K.R. Storage Stability of Solutions of DNA Standards. *Anal. Chem.* 2019, 91, 12268–12274.
- [74] Bonnet, J.; Colotte, M.; Coudy, D.; Couallier, V.; Portier, J.; Morin, B.; Tuffet, S. Chain and Conformation Stability of Solid-State DNA: Implications for Room Temperature Storage. *Nucleic Acids Research* 2010, 38, 1531–1546.
- [75] Deagle, B.E.; Eveson, J.P.; Jarman, S.N. Quantification of Damage in DNA Recovered from Highly Degraded Samples – a Case Study on DNA in Faeces. *Frontiers in Zoology* 2006, 3, 11.
- [76] Newman, S.; Stephenson, A.P.; Willsey, M.; Nguyen, B.H.; Takahashi, C.N.; Strauss, K.; Ceze, L. High Density DNA Data Storage Library via Dehydration with Digital Microfluidic Retrieval. *Nature Communications* 2019, 10, 1706.
- [77] van der Valk, T.; Pečnerová, P.; Diez-del-Molino, D.; Bergström, A.; Oppenheimer, J.; Hartmann, S.; Xenikoudakis, G.; Thomas, J.A.; Dehasque, M.; Sağlıcan, E.; et al. Million-Year-Old DNA Sheds Light on the Genomic History of Mammoths. *Nature* 2021, 591, 265–269.
- [78] Chatterjee, N.; Walker, G.C. Mechanisms of DNA Damage, Repair, and Mutagenesis. *Environmental and Molecular Mutagenesis* 2017, 58, 235–263.
- [79] Liu, X.; Jing, X.; Liu, P.; Pan, M.; Liu, Z.; Dai, X.; Lin, J.; Li, Q.; Wang, F.; Yang, S.; et al. DNA Framework-Encoded Mineralization of Calcium Phosphate. *Chem* 2020, 6, 472–485.
- [80] Paunescu, D.; Fuhrer, R.; Grass, R.N. Protection and Deprotection of DNA—High-Temperature Stability of Nucleic Acid Barcodes for Polymer Labeling. *Angewandte Chemie International Edition* 2013, 52, 4269–4272.
- [81] Puddu, M.; Paunescu, D.; Stark, W.J.; Grass, R.N. Magnetically Recoverable, Thermostable, Hydrophobic DNA/Silica Encapsulates and Their Application as Invisible Oil Tags. *ACS Nano* 2014, 8, 2677–2685.
- [82] Kohll, A.X.; Antkowiak, P.L.; Chen, W.D.; Nguyen, B.H.; Stark, W.J.; Ceze, L.; Strauss, K.; Grass, R.N. Stabilizing Synthetic DNA for Long-Term Data Storage with Earth Alkaline Salts. *Chem. Commun.* 2020, 56, 3613–3616.
- [83] Lau, B.; Chandak, S.; Roy, S.; Tatwawadi, K.; Wootters, M.; Weissman, T.; Ji, H.P. Magnetic DNA Random Access Memory with Nanopore Readouts and Exponentially-Scaled Combinatorial Addressing. *Scientific Reports* 2023, 13, 8514.
- [84] Sanger, F.; Nicklen, S.; Coulson, A.R. DNA Sequencing with Chain-Terminating Inhibitors. *Proceedings of the National Academy of Sciences* 1977, 74, 5463–5467.
- [85] Rodriguez, R.; Krishnan, Y. The Chemistry of Next-Generation Sequencing. *Nature Biotechnology* 2023.
- [86] Yeom, H.; Lee, Y.; Ryu, T.; Noh, J.; Lee, A.C.; Lee, H.-B.; Kang, E.; Song, S.W.; Kwon, S. Barcode-Free next-Generation Sequencing Error Validation for Ultra-Rare Variant Detection. *Nature Communications* 2019, 10, 977.
- [87] Yasumoto, S.; Muranaka, T. Foreign DNA Detection in Genome-Edited Potatoes by High-Throughput Sequencing. *Scientific Reports* 2023, 13, 12246.
- [88] Javed, N.; Farjoun, Y.; Fennell, T.J.; Epstein, C.B.; Bernstein, B.E.; Shores, N. Detecting Sample Swaps in Diverse NGS Data Types Using Linkage Disequilibrium. *Nature Communications* 2020, 11, 3697.
- [89] Teng, C.-F.; Huang, H.-Y.; Li, T.-C.; Shyu, W.-C.; Wu, H.-C.; Lin, C.-Y.; Su, I.-J.; Jeng, L.-B. A Next-Generation Sequencing-Based Platform for Quantitative Detection of Hepatitis B Virus Pre-S Mutants in Plasma of Hepatocellular Carcinoma Patients. *Scientific Reports* 2018, 8, 14816.
- [90] Chen, P.-C.; Yin, J.; Yu, H.-W.; Yuan, T.; Fernandez, M.; Yung, C.K.; Trinh, Q.M.; Peltekova, V.D.; Reid, J.G.; Tworog-Dube, E.; et al. Next-Generation Sequencing Identifies Rare Variants Associated with Noonan Syndrome. *Proceedings of the National Academy of Sciences* 2014, 111, 11473–11478.
- [91] de Masson, A.; O'Malley, J.T.; Elco, C.P.; Garcia, S.S.; Divito, S.J.; Lowry, E.L.; Tawa, M.; Fisher, D.C.; Devlin, P.M.; Teague, J.E.; et al. High-Throughput Sequencing of the T Cell Receptor β Gene Identifies Aggressive Early-Stage Mycosis Fungoides. *Science Translational Medicine* 2018, 10, eaar5894.
- [92] Clarke, J.; Wu, H.-C.; Jayasinghe, L.; Patel, A.; Reid, S.; Bayley, H. Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing. *Nature Nanotechnology* 2009, 4, 265–270.
- [93] Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; et al. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 2009, 323, 133–138.
- [94] Chin, C.-S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A.; et al. Phased Diploid Genome Assembly with Single-Molecule Real-Time Sequencing. *Nature Methods* 2016, 13, 1050–1054.
- [95] Flusberg, B.A.; Webster, D.R.; Lee, J.H.; Travers, K.J.; Olivares, E.C.; Clark, T.A.; Korlach, J.; Turner, S.W. Direct Detection of DNA Methylation during Single-Molecule, Real-Time Sequencing. *Nature Methods* 2010, 7, 461–465.
- [96] Ying, Y.-L.; Hu, Z.-L.; Zhang, S.; Qing, Y.; Fragasso, A.; Maglia, G.; Meller, A.; Bayley, H.; Dekker, C.; Long, Y.-T. Nanopore-Based Technologies beyond DNA Sequencing. *Nature Nanotechnology* 2022, 17, 1136–1146.
- [97] Koch, C.; Reilly-O'Donnell, B.; Gutierrez, R.; Lucarelli, C.; Ng, F.S.; Gorelik, J.; Ivanov, A.P.; Edell, J.B. Nanopore

- Sequencing of DNA-Barcoded Probes for Highly Multiplexed Detection of microRNA, Proteins and Small Biomarkers. *Nature Nanotechnology* 2023.
- [98] Plesa, C.; Verschuere, D.; Pud, S.; van der Torre, J.; Ruitenber, J.W.; Witteveen, M.J.; Jonsson, M.P.; Grosberg, A.Y.; Rabin, Y.; Dekker, C. Direct Observation of DNA Knots Using a Solid-State Nanopore. *Nature Nanotechnology* 2016, 11, 1093–1097.
- [99] Lopez, R.; Chen, Y.-J.; Dumas Ang, S.; Yekhanin, S.; Makarychev, K.; Racz, M.Z.; Seelig, G.; Strauss, K.; Ceze, L. DNA Assembly for Nanopore Data Storage Readout. *Nature Communications* 2019, 10, 2933.
- [100] Liu, Y.; Ren, J.; Qin, Y.; Li, J.; Liu, J.; Wang, E. An Aptamer-Based Keypad Lock System. *Chem. Commun.* 2012, 48, 802–804.
- [101] Meiser, L.C.; Gimpel, A.L.; Deshpande, T.; Libort, G.; Chen, W.D.; Heckel, R.; Nguyen, B.H.; Strauss, K.; Stark, W.J.; Grass, R.N. Information Decay and Enzymatic Information Recovery for DNA Data Storage. *Communications Biology* 2022, 5, 1117.
- [102] Purcell, O.; Wang, J.; Siuti, P.; Lu, T.K. Encryption and Steganography of Synthetic Gene Circuits. *Nature Communications* 2018, 9, 4942.
- [103] Zheng, L.L.; Li, J.Z.; Wen, M.; Xi, D.; Zhu, Y.; Wei, Q.; Zhang, X.-B.; Ke, G.; Xia, F.; Gao, Z.F. Enthalpy and Entropy Synergistic Regulation–Based Programmable DNA Motifs for Biosensing and Information Encryption. *Science Advances* 9, eadf5868.
- [104] Meiser, L.C.; Koch, J.; Antkowiak, P.L.; Stark, W.J.; Heckel, R.; Grass, R.N. DNA Synthesis for True Random Number Generation. *Nature Communications* 2020, 11, 5869.
- [105] Clelland, C.T.; Risca, V.; Bancroft, C. Hiding Messages in DNA Microdots. *Nature* 1999, 399, 533–534.
- [106] Zhang, Y.; Wang, F.; Chao, J.; Xie, M.; Liu, H.; Pan, M.; Kopperger, E.; Liu, X.; Li, Q.; Shi, J.; et al. DNA Origami Cryptography for Secure Communication. *Nature Communications* 2019, 10, 5469.
- [107] Grass, R.N.; Heckel, R.; Dessimoz, C.; Stark, W.J. Genomic Encryption of Digital Data Stored in Synthetic DNA. *Angewandte Chemie International Edition* 2020, 59, 8476–8480.
- [108] Zhang, Y.; Yin, X.; Cui, C.; He, K.; Wang, F.; Chao, J.; Li, T.; Zuo, X.; Li, A.; Wang, L.; et al. Prime Factorization via Localized Tile Assembly in a DNA Origami Framework. *Science Advances* 9, eadf8263.