



Review

DNA Data Storage

Tomasz Buko [†], Nella Tuczko [†] and Takao Ishikawa *

Department of Molecular Biology, Institute of Biochemistry, Faculty of Biology, University of Warsaw, Miecznikowa 1, PL-02-096 Warsaw, Poland

* Correspondence: t.ishikawa@uw.edu.pl

[†] These authors contributed equally to this work.

Abstract: The demand for data storage is growing at an unprecedented rate, and current methods are not sufficient to accommodate such rapid growth due to their cost, space requirements, and energy consumption. Therefore, there is a need for a new, long-lasting data storage medium with high capacity, high data density, and high durability against extreme conditions. DNA is one of the most promising next-generation data carriers, with a storage density of 10^{19} bits of data per cubic centimeter, and its three-dimensional structure makes it about eight orders of magnitude denser than other storage media. DNA amplification during PCR or replication during cell proliferation enables the quick and inexpensive copying of vast amounts of data. In addition, DNA can possibly endure millions of years if stored in optimal conditions and dehydrated, making it useful for data storage. Numerous space experiments on microorganisms have also proven their extraordinary durability in extreme conditions, which suggests that DNA could be a durable storage medium for data. Despite some remaining challenges, such as the need to refine methods for the fast and error-free synthesis of oligonucleotides, DNA is a promising candidate for future data storage.

Keywords: bit; byte; long term data storage; next-generation information storage; oligonucleotide; sequencing

Key Contribution: The latest achievements in DNA data storage are reviewed and summarized in simple way so that principles are understandable for biologists without a background in data science.



Citation: Buko, T.; Tuczko, N.; Ishikawa, T. DNA Data Storage. *BioTech* **2023**, *12*, 44. <https://doi.org/10.3390/biotech12020044>

Academic Editor: Marco Mesiti

Received: 20 April 2023

Revised: 22 May 2023

Accepted: 23 May 2023

Published: 1 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The demand for data storage is increasing by approximately 50% every year. In 2012, the entire world's total information storage was 2.7 ZB [1], in 2018 it reached 33 ZB, only to rise two-fold in 2020. It is estimated that newly created data will take up about 175 ZB by 2025 [2]. This equals a 65-fold increase only in the period between 2012 and 2025.

The tremendous Global Datasphere expansion is a strong motivator for new developments in data storage. Current data storage methods, such as magnetic (e.g., hard disk), optical (e.g., Blu-ray disc), and solid-state (e.g., flash drive), are insufficient to accommodate such rapid growth [3]. The main problems with those methods are their cost, space, and energy consumption during the recording, storing, and reading of data. Moreover, their durability reaches a maximum of 50 years in perfectly optimal conditions [4]. Humidity, extreme temperatures (both high or low), magnetic fields, or mechanical failures are the main reasons why those methods are not reliable for long-term data storage.

Therefore, there is a great demand for a new, longevous data storage medium with a high capacity, high data density, and high durability against extreme conditions [1]. There are a few prototypes of next-generation data carriers that may be able to cope with the above-mentioned challenges. Among them, DNA seems to be one of the most promising. The most distinguishing features of DNA from other storage media are its density and durability against the extreme conditions.

Escherichia coli has a storage density of 10^{19} bits of data per cubic centimeter [5]. This means that 1.7×10^{19} bits can be stored in just 1 g of DNA. Due to its three-dimensional structure, DNA is about eight orders of magnitude denser than other storage media. Moreover, DNA replication during PCR or the cell's proliferation enables the quick and inexpensive copying of vast amounts of data [3].

For years, a DNA specimen collected from a 700,000-year-old horse was considered to be the oldest extracted DNA. However, in 2021, this record was pushed to 1 million years. DNA extracted from mammoth teeth was successfully extracted and sequenced [6]. Additionally, scientists managed to sequence 300,000-year-old mitochondrial DNA from humans and bears [4]. These examples perfectly illustrate the longevity of DNA and proves its usefulness for archeological purposes or data storekeeping. If stored in optimal conditions and dehydrated, DNA can possibly endure for millions of years [1].

Numerous space experiments on microorganisms have proven their extraordinary durability in extreme conditions. Due to solar UV radiance, the space vacuum, and extreme temperature conditions, space is considered one of the most hostile environments [7].

UV radiance being the most deleterious parameter in space increases microorganisms' lethality by four orders of magnitude in relation to Earth's conditions [8]. UVB and UVC altogether cover the 200–315 nm light spectrum; these are the most hazardous to microorganisms and are responsible for their high lethality in space. This is caused by high irradiance absorption by DNA and proteins in such spectral ranges. In vegetative cells, this UV irradiance leads to DNA mutations, such as cyclobutane pyrimidine dimers and pyrimidine–pyrimidone photo products [9]. Meanwhile, in bacterial spores, thymine dimer photoproducts, so-called spore photoproducts (SP), are formed due to UV radiation. Despite this fact, all these dimers can be repaired by the direct reversal mechanism. Spores possess an additional SP-specific repair pathway that makes spores significantly more resistant to UV radiance than vegetative cells [10].

Regardless of such hostile conditions, it has been proven that spores of *Bacillus subtilis* shielded against UV solar radiation are able to survive in outer space for nearly 6 years. Although only 1–2% of the population recovered, the outcome was significantly increased (even to 90% of population recovered) if 5% glucose was added to the spore multilayer. It was suggested that glucose binds additional water molecules, preventing the cell from becoming completely desiccated. It also replaces water molecules, thereby stabilizing the macromolecular structure [8]. Furthermore, some microorganisms can even cope with a full space environment. For example, the lichens *Rhizocarpon geographicum* and *Xanthoria elegans* survived a 2-week exposure to outer space. After that time, the lichens completely restored their photosynthetic activity and no ultrastructural changes were revealed in most of the fungal and algal cells of lichens [11]. It is supposed that their thick cortex with UV-screening pigments (rhizocarpic and parietin phenolic acids) are responsible for their survival [12].

2. Coding Files in DNA

Encoding information in DNA is based on binary code. A specific nucleotide corresponds to a code, for example, 00 → A, 01 → C, 10 → G, and 11 → T. While binary data are “translated” into a DNA sequence, it is important to avoid long homopolymers (more than three same nucleotides in a row) and unreasonable GC content, as both might generate mistakes during the synthesis and sequencing of DNA strings. In fact, encoding a file requires converting text into a code such as ASCII (Figure 1) or Base64, and then converting the coded file into a binary system. The encoding field uses different coding algorithms, such as Huffman, to condense messages and balance code, preventing homopolymer sequences. Two examples of coding systems, their modifications, and other algorithms of a similar kind generate proper DNA strings [13,14], which are capable of long data storage.

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

Figure 1. An example of coding the message “ramy” into an ASCII code. Converting binary data into nucleotide sequences is made by computer algorithms.

Church et al. (2012), for the first time, encoded a draft of a book, eleven JPG images and one JavaScript program in DNA [15]. For this purpose, they used a simple encoding method involving the translation of zeros into A or C and ones into T or G. As a result, the authors received 54,898 oligonucleotides, each containing three parts: 96 bases of data, 22-bases-long sequences at both ends, allowing those oligonucleotides to be parallelly amplified by PCR, and the 19-bases-long index sequence, pointing out the segment position in the original file [15]. Encoding one bit per base allowed the authors to avoid sequences that were potentially hard to write or read. Splitting information into blocks of data allowed the authors to circumvent the problems associated with the synthesis of long DNA strings. This pioneering work demonstrated the real possibility of using DNA as a data storage material, and also showed the enormous capacity of this method. An important element of the works of that time was to show the limitations of the method used. Through this work, it was noted that the information encoded in DNA is prone to sequencing errors, mainly in homopolymer regions.

One year later, Goldman et al. (2013) tried to overcome the sequencing errors occurring by encoding data with redundancy [16]. The authors encoded all 154 of Shakespeare’s sonnets, a scientific article, a medium-resolution color photograph of the European Bioinformatics Institute, and a 26 s long excerpt from Martin Luther King’s 1963 “I have a dream” speech using the Huffman algorithm to covert numeric data into a nucleotide sequence [16]. In summary, bytes of binary sequences were converted into base-3 digits (or ternary) from 0 to 2, which were then associated with three nucleotides, A, T, and C (or G if C has been used for the encoding of the previous ternary digit). DNA strings were divided into 100-nucleotide-long oligos with an overlap of 75 residues between adjacent fragments, creating four-fold redundancy (Figure 2). Alternate fragments were converted to their reverse complement, which reduces the probability of systematic failure, such as issues with DNA sequencing. Indexing sequences comprising 17 nucleotides were also encoded at the beginning and end of each fragment.

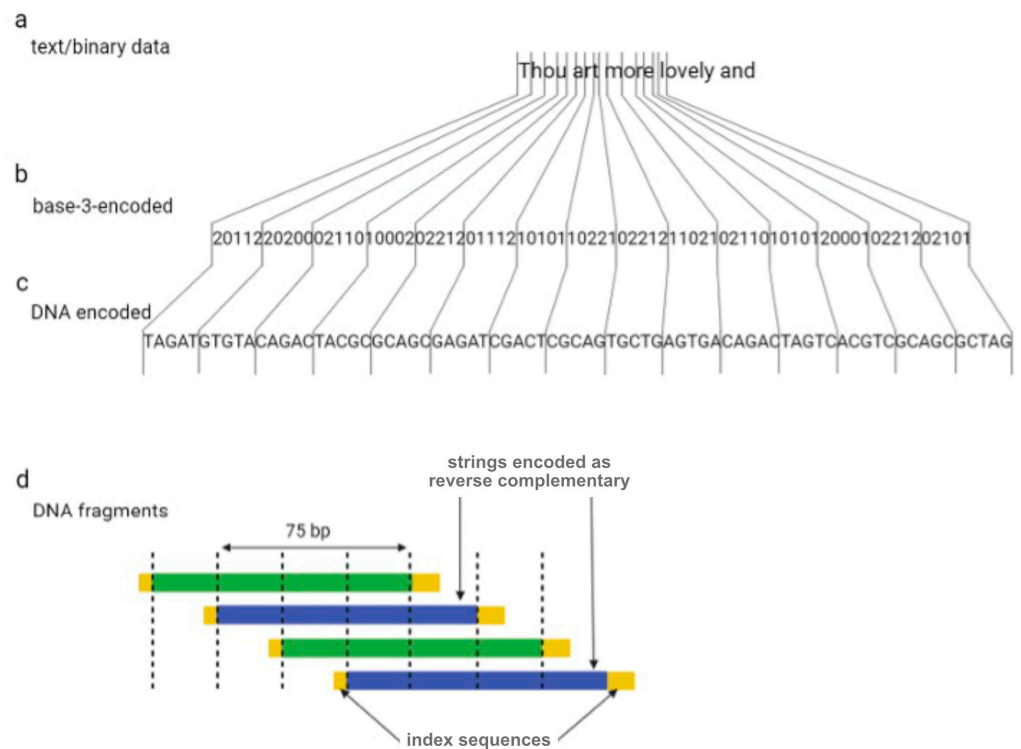


Figure 2. The coding scheme implemented by Goldman et al. Digital information (a) is converted to base-3 (b) using a Huffman code and is subsequently is converted to DNA strings (c). Dividing DNA strings as shown generated four-fold redundancy (d).

Ailenberg and Rotstein (2009) encoded text, music, and images in DNA by using modified Huffman coding (Figure 3) [17]. In their work, they constructed a plasmids library each containing 10,000 bp of information and an index plasmid that contains basic information, such as the title, author, plasmid number, and primer assignments used to read coded information [17]. The authors also constructed a separate encoding table for each type of file, which allowed the authors to encode each character from the keyboard. The authors also indicated the possibility of extending their code according to the described rules.

a

	note values	note pitches	meter	repeat
G	Quarter note (1/4)	TG A	CAT 2/4	CAG (
TT	Half note (1/2)	AC B	CAA 3/4	CCA)
TA	Whole nota (1)	AG C	CAC 4/4	CCT ×
AT	Eighth note (1/8)	CG D		CCG 2
CT	Sixteenth note (1/16)	AAT E		CCC 3
TC	dot (.)	AAC F		AAAT 4
		AAG G		

b

file type		Ma-	ry	had	a	lit-	tle	lamb	lit-	tle	lamb,	lit-	tle	lamb			
m	u	*	4/4	(B1/4	A1/4	G1/4	A1/4	B1/4	B1/4	B1/2	A1/4	A1/4	A1/2	B1/4	D1/4	D1/2
taac tact tacca cac cag ACG TGG AAGG TGG ACG ACG ACTT TGG TGG TGTT ACG CGG CGTT																	
Ma- ry had a lit- tle lamb, its fleece was white as snow																	
B1/4 A1/4 G1/4 A1/4 B1/4 B1/4 B1/4 B1/4 A1/4 A1/4 B1/4 A1/4 G1) × 2																	
ACG TGG AAGG TGG ACG ACG ACG ACG TGG TGG ACG TGG AAGTA cca cct ccg																	

Figure 3. An example of coding music in DNA. Fragment of “Mary Had a Little Lamb” encoded using Huffman code. A nucleotide sequence corresponding to the music code is shown in (a) and the encryption part in (b). Adapted from Ailenberg and Rotstein [17].

The first example of the graphical file recoded in DNA was a simplified lamb drawing (Figure 4). Although this image consists of simple geometric figures, the simplicity and geometry of the image are not general requirements. Yazdi et al. (2017) managed to encode The Citizen Kane poster photograph and Smiley Face emoji (Figure 5) [18]. For this purpose, they used Base64 encoding to convert files into binary format. The DNA string length used by the authors was 1000 bp, containing 984 bp of information and 16 bp of address sequence. The purpose of the addressing method was to enable random access to codewords via highly selective PCR reactions. This approach allows the specific amplification of a pool of oligos without amplifying and reading all sequences from a given pool. This work also presented a new deletion-correcting method called homopolymer check codes. This method of correction divides DNA sequences into strings of homopolymers, e.g., {AATCCCCGA} into strings {AA, T, CCC, G, A}, which gives a homopolymer sequence of length {2,1,3,1,1}. The homopolymer length sequence contains special redundancy that protects against asymmetric substitution errors. Hypothetically, when two deletions occur in the sequence resulting in {ATCCGA}, the length of the homopolymer fragments is {1,1,2,1,1}. Recovering the original sequence is possible by correcting two bounded magnitude errors. Combining this with GC content balancing, the subsequent alignment of DNA oligonucleotides, and post-sequencing sequence sorting based on the correctness of the index sequence resulted in a new coding method.

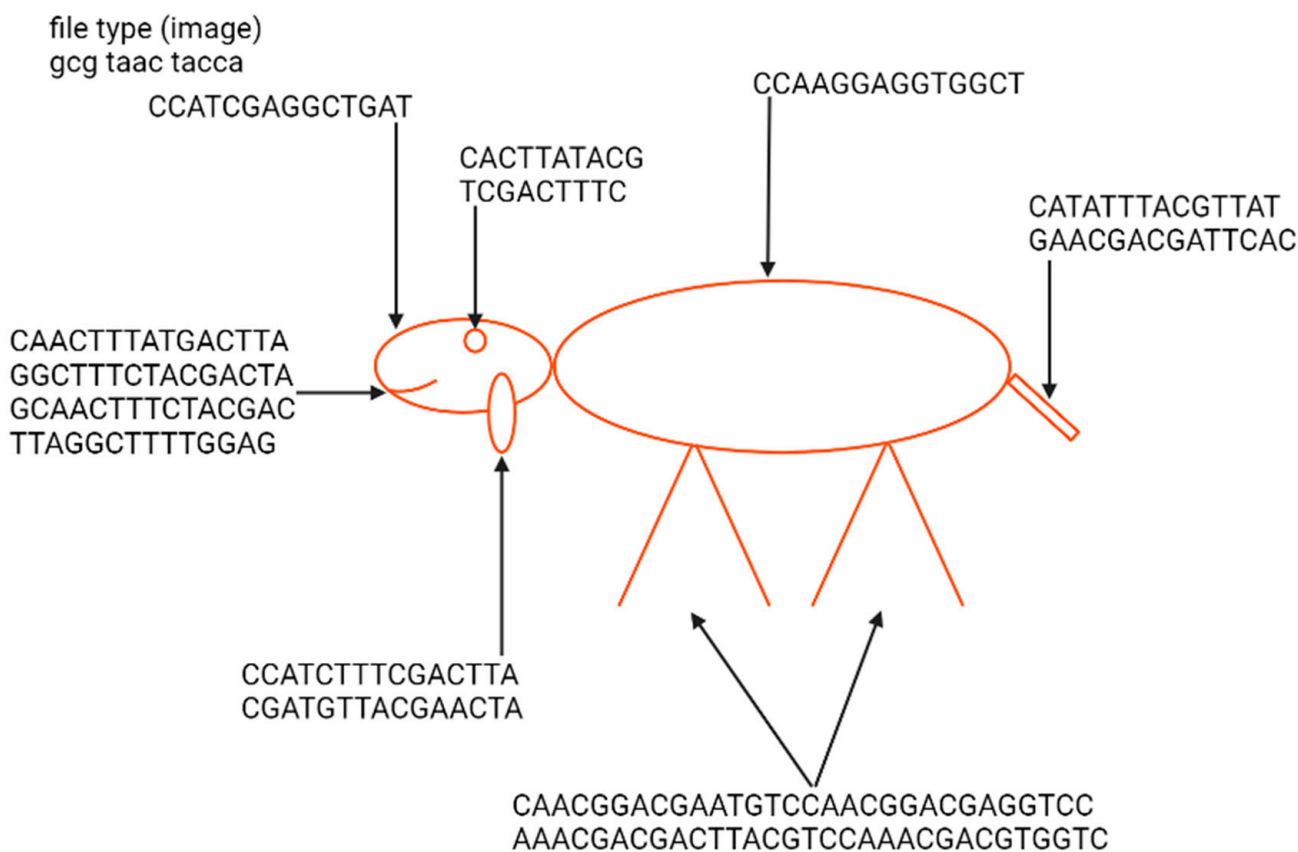


Figure 4. Indication of elements of the nucleotide sequence in which a Little Lamb was encoded and an example image presenting a lamb from the “Mary Had a Little Lamb” rhyme encoded by Ailenberg and Rotstein [17]. The sequence of a file type defines it as an image. The geometric shape of the lamb enables the use of only 238 bp of DNA for encoding. Encoding has been performed using a template of signs indicating the type of shape and its spatial coordinates.



Figure 5. Smiling emoji and original Citizen Kane poster photograph encoded and decoded by Yazdi et al. [18]. The raw images were encoded and synthesized in the form of DNA strings (a,b). Images received after decoding without homopolymer check codes during processing (c,d). Images received after sequencing DNA strings when homopolymer error correction was made in order to reduce the number of errors that occurred during each encoding and decoding step (e,f). Two errors in the Citizen Kane file were sufficient to make the recovery of the image impossible. One error in the emoji did not influence the image quality.

Coding motion picture as motion GIFs and movies has also been achieved in the DNA data storage field. In 2017, Shipman et al. encoded five frames of a galloping mare from Eadweard Muybridge’s “The Human and Animal Locomotion Photographs” [19]. In their experiment, CRISPR-Cas was used to integrate an encoded short movie into the genomes of a population of living bacteria. The usage of this method does not change the overall encoding protocol. Strings of DNA are integrated into the CRISPR array thanks to appropriate integrases. Spacer sequences in the CRISPR array were used to encode barcodes defining which set of pixels was encoded in a specific part. The use of the CRISPR method for GIF encoding was of great importance because it allows the encoding of subsequent sequences without the need to additionally index them. This is because newly added sequences are almost always integrated in such a way that they push the previously integrated sequences away from the leader region. Therefore, the order of the sequence was conditioned by successive transformations in which DNA with encoded movie frames was introduced to bacterial cells.

A number of other works referring to information encoding in DNA are summarized in Table 1 below.

Table 1. Works regarding the coding of information on DNA. In “redundancy or error correction” column, “n.d.” indicates that there is no information in the original work.

Authors	Data Size	Length of Strings	Encoding Method	Redundancy or Error Correction	Modification	Reference
Bornholt et al.	51 KB	120	Huffman code	DNA string exclusive-or	–	[20]
Blawat et al.	22 MB	230	Own bit mapping	BCH code	–	[21]
Organick et al.	200 MB	~150	Base-4	Reed–Solomon	–	[22]
Choi et al.	854 B	85	Own bit mapping	Reed–Solomon	Degenerate bases	[23]
Lee et al.	96 B	~50	ASCII	codec	Enzymatic DNA synthesis	[24]
Tabatabaei et al.	2 KB; 392 KB	450	Own bit mapping	Not needed	Enzymatic nicking (<i>Pf</i> Ago)	[25]

Table 1. Cont.

Authors	Data Size	Length of Strings	Encoding Method	Redundancy or Error Correction	Modification	Reference
Yang et al.	23 KB	83	A, C = 0; G, T = 1	n.d.	TNA	[26]
Ren et al.	682 B; 39 KB; 28 MB	~100	RABR; RALR	Reed–Solomon	Artificial nucleotides	[27]
Mayer et al.	24,5–33,6 KB	~40	ASCII; Elias gamma	n.d.	Epigenetic encoding	[28]

3. Synthesis of DNA Strings

Chemical DNA synthesis has made tremendous progress since the 1970s, when fragments of about 20 nucleotides could be synthesized, to the present, when fragments of up to 500 nucleotides can be easily made. The technology commonly used for the synthesis of DNA strands enables only short 200–300 nucleotides sequences to be synthesized, which is a limitation when coding a large amount of data. Nevertheless, the technology used for DNA synthesis on microarrays seems to be more suitable for this purpose. It allows the synthesis of parallel oligonucleotides containing different sequences (Figure 6). By using it, the time and cost needed for the synthesis of large-scale DNA libraries might be greatly reduced [29]. Microarrays have enabled the high-fidelity synthesis of oligo pools of about 300 nucleotides in length [30]. Regardless of the synthesis method, long DNA fragments must be assembled from oligos. It is also necessary to add indexes to each fragment, or sequence overlapping in successive DNA fragments [3], unless—as discussed above—the CRISPR method is used to record information in the bacterial genome. In 2017, Heckel et al. considered the storage capacity using both assembly methods and have shown that an index-based coding system is optimal for data storage purposes [31].

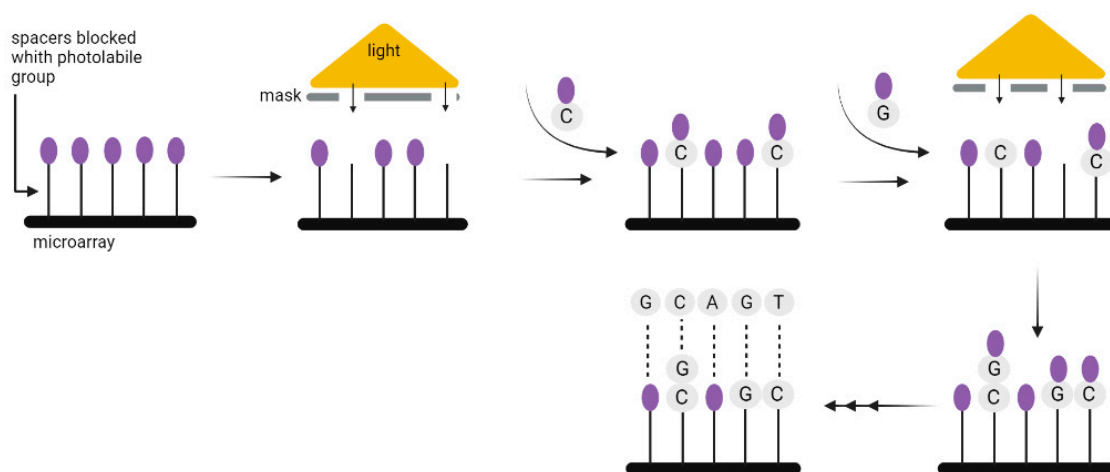


Figure 6. A solid-phase method for the synthesis of oligonucleotides using photolabile compounds. A spacer containing the photolabile group is covalently joined to the surface. Once spots on the surface are exposed to UV light through slits in the physical mask, the photolabile protecting group is removed and the synthesis of oligonucleotide begins. The subsequent appropriate phosphoramidite with the photolabile group is then applied to the entire surface of the plate. It can form covalent bonds only in the absence of the preceding photolabile group. In the subsequent steps, additional spots are exposed to radiation, and another phosphoramidite is applied where necessary. Until the final oligonucleotide is completely synthesized, the chain-extending processes are repeated [29].

4. New Storage Medium, Old Problems, and Solutions

A serious problem with the usage of DNA for data storage purposes is that long-term storage, synthesis, and sequencing might introduce some errors (such as deletion, insertion,

or substitution). It should be stressed that errors are not the only issue when DNA is used as the data storage medium, but this is a problem of all information storage technologies. This is why there is a solution to it in the form of error-correcting codes (ECCs), in which a minimal amount of special data is added for error-correction purposes. In classical data-storage devices, the use of ECCs adds redundancy and allows the correction of essentially all errors that occur during use. ECCs such as fountain code, rapid tornado code, HEDGES (Hash Encoded, Decoded by Greedy Exhaustive Search), or the Reed–Solomon code [32] are used in DNA data storage. In general, ECCs introduce sequence redundancy, which enables the subsequent recovery of complete data even in the case that some oligonucleotides used for data storage are physically damaged. The implementation of ECCs slightly diminishes the storage capacity (because ECCs are often based on adding external fragments to the sequences encoding data), but its advantages—namely the possibility of error correction—outweigh this limitation. ECCs enable insertions and deletions to be corrected, as well as the loss of some parts of the DNA strings. An alternative to ECCs was the previously used high-depth sequencing, which, for obvious reasons, only corrected sequencing errors.

One of the most frequently mentioned ECCs in the literature is a Reed–Solomon code (Figure 7). In general, the Reed–Solomon code is based on the transformation of the original data set to a symbol set. The symbols are then converted to coefficients in a system of linear equations and their solutions enable the original data set to be accessed. Meiser et al. (2020) have used a Reed–Solomon code for storing a full album of music in DNA [33].

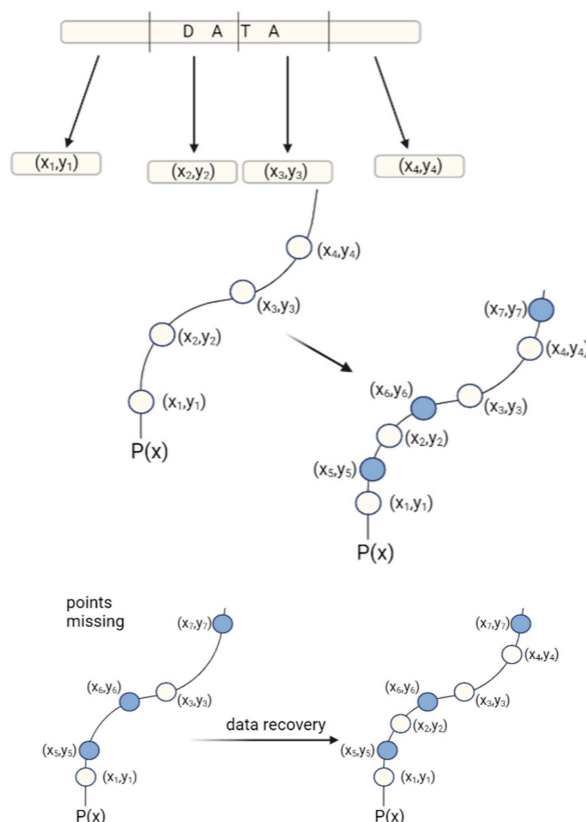


Figure 7. Principle of Reed–Solomon correction: first, the data is divided into parts, and each part is assigned x and y values that determine its location. Based on the coordinates, the points are matched to the polynomial function $P(x)$, which is used to determine the parity symbols. Parity symbols are extra data points that match the original DNA sequence and are stored with the original data. When some of the original data are lost, the remaining data points and parity symbols can be used to recreate the original polynomial function and receive original data.

Recently, Xie et al. (2023) conducted an analysis showing the value of the sequencing depth for retrieving the right string of data [34]. Sufficiently deep sequencing allows the

use of MSA (multiple sequence alignment) methods to establish a consensus sequence and correct errors that may appear on the DNA strands. The MAFFT algorithm was chosen for the analysis, which has been shown to be able to correct more than 95% of errors at a sequencing depth reaching $100\times$ when the error rate is lower than 15%. The authors showed that adequately deep sequencing combined with MSA is able to correct errors when their frequency is less than 20%. Above this value, error correction based on MSA is possible with the simultaneous use of ECC. This method enables the cost and time reduction needed for the DNA data storage procedure.

Erlich and Zielinski (2017) used the fountain algorithm to encode 2.14×10^6 bytes of data [35]. The fountain encoding algorithm works in three steps: preprocessing, the Luby transform, and screening (Figure 8). Overall, it aims to convert the input file into a collection of DNA strings that pass synthesis and reading constraints.

Preprocessing—In this step, the input file is compressed using a lossless algorithm. Then, the algorithm partitions the file into non-overlapping K segments, in which each segment is L bits long. L is defined by the user.

Luby transformation—This step consists of many substeps. Briefly, a pseudo-random number generator determines the number of segments that will be packed into a single packet. Encoded segments become packets known as droplets. For this, the algorithm uses a robust solution probability distribution, which assumes that most of the droplets will be created with a small number of input segments. On the segments of one droplet, the algorithm performs a bitwise exclusive or XOR operation. For example, consider that the algorithm randomly selected three input fragments: 0100, 1100, 1001. In this case, the droplet is $0100 \oplus 1100 \oplus 1001 = 0001$. In the end, the algorithm adds an index that specifies the binary representation of the seed, which, in turn, corresponds to the state of the random number generator of the transform during the generation of the droplet. Finally, it enables the decoder algorithm to infer the identities of the segments in the droplet.

Screening—In the last step, the algorithm excludes those strings that do not pass the biochemical constraints. Firstly, binary data are translated into a nucleotide sequence: {00, 01, 10, 11} to {A, C, G, T}. Then, DNA strings are screened for GC content and homopolymers. The sequences that do not pass the screen are removed and the formation and screening of the oligonucleotides are repeated until the desired conditions are obtained. In practice, the authors recommend synthesizing 5–10% more oligonucleotides than the input segments.

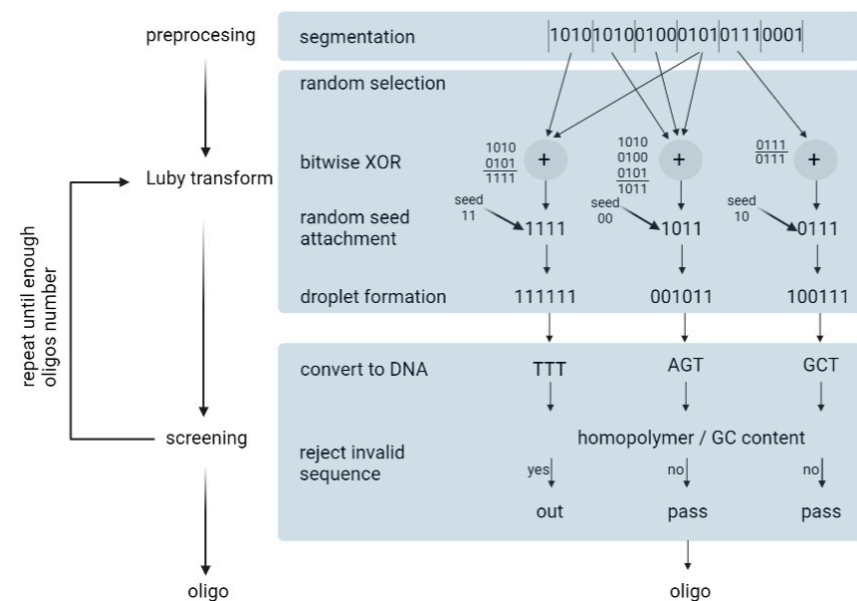


Figure 8. Depiction of DNA fountain strategy.

The idea for the decoding algorithm is to start with single-segment droplets and propagate that information through the other droplets until all the segments are recovered.

5. DNA Preservation

Although the theoretical density of DNA data storage reaches petabytes per gram, usually this value is unreachable. Due to the necessity of adding protective substances to the DNA, the loading efficiency (DNA weight/total weight) ranks below 100%. Moreover, the presence of indexes, such as Reed–Solomon codes, in long strands of DNA cause the loss of data storage density. It was estimated that the index ratio of 200 bp DNA reaches 6.5%. Furthermore, DNA without protection is liable to degradation due to physical and chemical factors, such as temperature, water, UV irradiation, oxidation, or extreme pH values [36]. Therefore, current research focuses on increasing the DNA data storage density and the time of its preservation by protecting DNA from the influence of high humidity and the presence of oxygen [37].

The methods used for DNA preservation can be divided into two essential categories: *in vitro* preservation, where DNA is usually stored in a single physical DNA pool, or *in vivo* preservation, which uses living cells as DNA carrier systems [32].

5.1. *In Vitro* Preservation

The most common way to store data within DNA *in vitro* is solution storage. At first, DNA was preserved in ethanol, however, over time the ammonium-based ionic liquids gained popularity. Due to hydrogen bonding between ionic liquid and DNA, those solutions improve DNA stability. However, the solution storage allows DNA to be stored for only a year, which is insufficient to fulfill the aims of DNA data preservation (>1000 years).

On the contrary, solid-state DNA appears to be more stable due to its reduced molecular mobility and lack of water, which causes hydrolytic damage [35]. The successful amplification of DNA from ancient specimens, such as the Pleistocene cave bear, additionally indicates the effectiveness of the method [37]. Based on this discovery, Grass and co-workers proposed DNA silica fossilization technology, through which they obtained stable DNA after 35 days in 65 °C (equivalent to two years at room temperature) [38]. Furthermore, Newman et al. (2019) developed a method for the preservation of dehydrated DNA spots on glass cartridges, which can subsequently be recovered by a water droplet. Multiple DNA spots on one cartridge additionally increase the storage density of 50 TB of data per glass cartridge [39]. Choi et al. (2020) created a DNA micro-disc, which allows easy access to data-encoded DNA and write-once-read-many memory. Firstly, the encoded DNA's primer sequences and data description were included in the QR code, which facilitates easy access to the data. Secondly, due to the immobilization of DNA on the micro-disc, after DNA enrichment using PCR, the original and amplified DNA are separated. The sequence of the amplified DNA is subsequently converted into binary data and the immobilized DNA can be read out in the future. Eventually, Choi et al. (2020) reached a density of up to 10^{12} bit/mm³ for a single micro-disc and assessed the durability of dehydrated DNA over 100 years at a temperature below 10 °C [40].

DNA can also be easily stored via freeze drying or the addition of additives. In fact, the lower the temperature, the longer the possible preservation. However, lyophilization may cause cytolysis due to the formation of ice cracks [36]. Moreover, the estimated annual cost of maintaining frozen samples around the globe likely surpasses USD 100 million each year [41]. Therefore, due to the high cost currently, scientists are trying to develop an effective method of DNA preservation at room temperature. For instance, the addition of additives such as trehalose or PVA enables the DNA to be preserved at room temperature. Both stabilizers create hydrogen bonds with negatively charged phosphate groups in DNA, which has a protective effect on its stability [36]. However, Ivanowa and Kuzmina (2013) indicate that, generally, the additives are insufficient for long-term DNA storage. Diluted DNA in trehalose solution stored for a month at room temperature granted only 46% PCR

success, and 2-year preservation in Tris-buffered PVA granted 50% PCR success, where PCR success was calculated as a percentage of positive wells per plate (96 samples) [42].

In Table 2, we summarize the storage methods used and the PCR success after storage for a specified period at a specified temperature.

Table 2. Storage methods of DNA and PCR success after the storage.

Storage Method	Time	Temperature	PCR Success	Reference
Chemical encapsulation				
Silica nanoparticles	9 months	RT	x	[43]
DNA-layered titanate nanohybrid	1 month	x	x	[44]
Solution Preservation				
“DNA stable”	4 years	RT	98%	[42]
DMSO salt solution	4 months	RT	42%	[45]
DMSO salt solution	2 years	RT	x	[46]
70% ethanol	4 months	RT	27%	[45]
70% ethanol	2 years	RT	x	[46]
90% ethanol	6 months	RT	96%	[47]
Formalin-fixed	30 years	RT	30%	[48]
Formalin-fixed	2–6 years	RT	x	[49]
Paraffin-embedded tissues	2–6 years	RT	x	[49]
DETs buffer	6 months	RT	92%	[47]
TE buffer	1 night	−20 °C	100%	[50]
TE buffer	3 years	−20 °C	x	[51]
Dehydration				
Ancient bone	521 years	13 °C	x	[52]
Filter Paper	4 years	RT	82.5%	[53]
Dried DNA	4 months	RT	35%	[45]
FTA cards	up to 128 days	RT	95%	[54]
Silica Gel	6 months	RT	50%	[47]
Oven-dried	6 months	RT	72%	[47]
Oven-dried	6 months	−20 °C	86%	[47]
Freeze drying				
DNA	4 years	4 °C	49%	[42]

RT is abbreviation for “room temperature”. X indicates that the information was not specified in the reference.

In Table 3, we present the durability of DNA in various accelerated aging tests. Such tests are performed to simulate the long-term behavior of DNA molecules in a much shorter time by applying harsh conditions. The results of those experiments are presented as C/C₀ (%), which is the percentage of the initial amount of DNA present in the sample after the accelerated aging test.

Table 3. The durability of DNA in accelerated aging tests.

Storage Method	Time	Temperature	Relative Humidity	Half-Life	Temperature	C/C ₀	Reference
Experimental Conditions				Parameters in Non-Experimental Conditions			
Chemical encapsulation							
Silica nanoparticles	2 weeks	70 °C	50%	20–90 years	20 °C	90%	[43]
Silica nanoparticles	10 days	60 °C	50%	5 months	RT	65%	[55]

Table 3. Cont.

Storage Method	Time	Temperature	Relative Humidity	Half-Life	Temperature	C/C ₀	Reference
Experimental Conditions				Parameters in Non-Experimental Conditions			
Calcium phosphate crystals	6 days	70 °C	50%	1 year	10 °C	0.1%	[56]
"DNAshell"	2 days	100 °C	50%	1 million years	25 °C	x	[57]
"DNAshell"	30 h	76 °C	50%	100 years	25 °C	x	[58]
"DNAshell" + trehalose	1 month	76 °C	50%	2000 years	25 °C	x	[58]
In silica	1 week	70 °C	50%	200 years	10 °C	10%	[4]
Solution Preservation							
"DNA stable"	1 week	65 °C	50%	4 years	25 °C	10%	[4]
"GenTra"	1 week	65 °C	50%	2 years	25 °C	50%	[59]
TE buffer	20 days	65 °C	50%	20 years	−20 °C	x	[51]
Dehydration							
DNA	6 weeks	50 °C	50%	x	x	10%	[60]
DNA silica fossilization	35 days	65 °C	50%	2 years	RT	15%	[61]
Dehydration with earth alkaline salts	6 days	70 °C	50%	750 years	10 °C	10%	[62]
DNA micro-disc	2 weeks	70 °C	50%	>700 years	0 °C	x	[40]
DNA with trehalose	10 days	70 °C	75%	17 years	10 °C	x	[63]
Filter card	1 week	70 °C	50%	3.7 years	25 °C	1%	[4]
Freeze drying							
Polymer-plasmid complexes	10 months	40 °C	50%	3 years	RT	x	[64]
Trehalose	2 months	60 °C	50%	2 years	RT	x	[65]
Cryosilicified samples	4 weeks	70 °C	60%	1200 years	20 °C	31%	[66]
Additives							
Trehalose	2 years	56 °C	50%	20 years	RT	50%	[42]
Trehalose	1 week	65 °C	50%	160 years	10 °C	20%	[59]
PVA	2 years	56 °C	50%	20 years	RT	15%	[42]
"Sugar mix"	1 week	65 °C	50%	1 year	20 °C	30%	[59]

RT is abbreviation for "room temperature". x indicates that the information was not specified in the reference.

5.2. In Vivo Preservation

Recently, in vivo preservation has been intensively developed. Preservation within a living cell allows the DNA to be replicated with a few orders of magnitude, much faster than by PCR, during the cell's proliferation processes [67].

Bacteria are the most intuitive way to preserve DNA within a living organism. However, during bacterial replication, the spontaneous mutation rate is 2.2×10^{-10} mutations per nucleotide per generation, or 1.0×10^{-3} mutations per genome per generation [68]. A generation time of about 20–30 min for *E. coli* means that after a few years of cultivation, mutations might represent a significant problem. Furthermore, the size of the introduced plasmid is a serious limitation of in vivo preservation methods. So far, the greatest amount of information in vivo has been encoded by Hao et al. (2020) thanks to the mixed-circle method developed by them. The procedure involves the cloning of data-encoded DNA oligonucleotides into plasmids and transforming *E. coli* cells with recombinant, data-containing plasmids. During data recovery, plasmids are sequenced, and oligonucleotides are assembled into original sequence. Eventually, 2304 kbp synthetic

oligonucleotides (encoding 455 KB of digital files) were used to create the mixed culture of bacterial cells [67].

The solution to the problem of the limited size of the introduced plasmid appears to be in vivo preservation on a yeast artificial chromosome. In 2021, Chen et al. created a circular 255 kbp yeast artificial chromosome (a data-carrying chromosome; dChr) encoding a total of 38 KB of digital data (two pictures and a video) [69]. Moreover, the dChr was replicated with high fidelity, no mutation appeared after the 100th generation of replication, while the encoding method used in this setup was tolerant toward a comparatively low accuracy of Nanopore sequencing, enabling the fast retrieval of reliable data [69]. The high fidelity of dChr replication could be achieved due to its chromatin-like structure formed in vivo [70]. As it is known that nucleosomes regulate DNA repair mechanisms [71,72], the utilization of eukaryotic organisms, such as *Saccharomyces cerevisiae*, carrying dChr is one of the promising approaches for DNA data storage.

Another approach to in vivo storage is the preservation of data in endogenous DNA, such as genomic DNA. This can be achieved using DNA-modifying enzymes such as nucleases, integrases, or recombinases, although recently, the CRISPR-Cas9 system has gained much popularity [73]. At the beginning of 2022, Liu et al. used a dual-plasmid system based on a single crRNA-guided endonuclease (CRISPR-Cas12a) to encode a codebook (56 bytes) and a picture (376 bytes) [74]. The authors used two plasmids, one with data-encoded (target) DNA and the second with templates for the expression of Cas protein and crRNA, which after bacteria transformation, enabled the introduction of target DNA to the *E. coli* genome. Ultimately, the rewriting reliability reached 94% and the information sequenced from the 252nd generation was 100% correct [74].

Studies on antimutator phenotypes have provided valuable insights into the sources and mechanisms of spontaneous mutations. Research on carbon-starved *E. coli* populations has shown that stress responses are required for the mutagenic repair of DNA breaks [75]. In the growing *E. coli* population, mutants of the α subunit of replicative DNA polymerase III have been well characterized as antimutator alleles, suggesting that DNA replication errors are a major source of spontaneous mutagenesis under optimal growth conditions [76]. However, these alleles also reduce specific transition mutations, making it unclear whether replication errors in wild-type cells stem from the intrinsic fidelity of DNA polymerase III or specific subpopulations with unique properties [77].

Despite the understanding of the molecular mechanisms controlling mutagenesis, the process of spontaneous mutation in cells with functional mutation-prevention systems remains unknown. To investigate this, a mutation assay on isogenic *E. coli* cells growing optimally without external stress was performed. It was revealed that spontaneous DNA replication errors occurred more frequently in subpopulations experiencing internal stresses, such as issues with proteostasis, genome maintenance, and reactive oxidative species production. These mutator subpopulations do not significantly impact the average mutation frequency or the overall fitness of the population in a stable environment. However, they play a crucial role in enhancing population adaptability in fluctuating environments by providing a reservoir of increased genetic variability [78].

In turn, such mutator subpopulations may be responsible for introducing spontaneous mutations in the *E. coli* population used for DNA data storage. Further understanding the molecular background of spontaneous mutations may be helpful in minimizing the occurrence of errors in the DNA used as a data storage medium in in vivo preservation methods.

6. DNA Sequencing

To convert the DNA sequence back to its digital code, DNA has to be sequenced and decoded to digital data using computer algorithms. Currently, the most commonly used platforms for the sequencing of data-encoding DNA are Next-Generation Sequencing by Illumina sequencing and Third Generation Sequencing by Oxford Nanopore Technology [37].

One of the biggest advantages of Nanopore over Illumina for data output purposes is its single-molecule sequencing of the extended alphabet, or its ability to sequence not only natural nucleotides, but also chemically modified nucleotides. The applicability of such an extended alphabet could significantly improve data storage in DNA by increasing storage density and, possibly, writing speed [79]. However, Nanopore also has some limitations, for instance, lower accuracy compared to Illumina. In fact, a direct comparison of the error rates of Nanopore (~10% per nucleotide in single read-out) and of Illumina (~0.5% per nucleotide) shows that Nanopore technology is approximately 20 times less accurate. Therefore, at the moment, for DNA data storage purposes, the most commonly used is Illumina sequencing [37].

7. Conclusions

Modern societies generate huge amounts of data and the rate of their growth has multiplied in recent years. The need to store both currently generated data and those generated in the past using classical data storage methods are consuming huge financial outlay and physical space. It also entails high costs for the environment, with the introduction of new methods of data storage thus urgently required.

For a long time, people have paid attention to the high storage density and longevity of DNA. In this article, we have provided a brief overview of how information is encoded and stored in DNA. The continuous development of these methods leads to a reduction in the number of errors appearing in the encoding and decoding processes, extending the durability of DNA as a data carrier, and reducing the cost of its storage.

Despite the continued growth in the field of information storage on DNA, some challenges still remain. There is a need to refine the methods used for the fast and error-free synthesis of oligonucleotides, and in the long run, also of long DNA chains. The method used to read nucleotide sequences also must evolve towards greater credibility.

Despite the current obstacles, the prospects for implementing data storage on DNA are very promising. There are even new ideas related to the use of chemical analogues of DNA, such as TNA, with even higher possible storage densities [26].

Author Contributions: Conceptualization, T.B., N.T. and T.I.; supervision, T.I.; writing—original draft, T.B. and N.T.; writing—review and editing, T.B., N.T. and T.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Figure 5 originates from: Yazdi et al., Portable and Error-Free DNA-Based Data Storage. *Sci. Rep.* 2017, 7, 5011, Springer Nature, distributed under the Creative Commons Attribution 4.0 International License. We changed emoji pictures in panels: b, d, and f. Figure 6 is based on the Figure 1 from: Sinyakov et al., Application of Array-Based Oligonucleotides for Synthesis of Genetic Designs. *Mol. Biol.* 2021, 55, 487–500, Springer Nature. It has been reproduced with permission from Springer Nature.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. De Silva, P.Y.; Ganegoda, G.U. New Trends of Digital Data Storage in DNA. *BioMed Res. Int.* **2016**, *2016*, 8072463. [[CrossRef](#)] [[PubMed](#)]
2. Rydning, J.; Reinsel, D.; Gantz, J. *The Digitization of the World from Edge to Core*; IDC: Framingham, MA, USA, 2018.
3. Ceze, L.; Nivala, J.; Strauss, K. Molecular Digital Data Storage Using DNA. *Nat. Rev. Genet.* **2019**, *20*, 456–466. [[CrossRef](#)] [[PubMed](#)]
4. 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. *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 2552–2555. [[CrossRef](#)] [[PubMed](#)]

5. Zhirnov, V.; Zadegan, R.M.; Sandhu, G.S.; Church, G.M.; Hughes, W.L. Nucleic Acid Memory. *Nat. Mater.* **2016**, *15*, 366–370. [[CrossRef](#)] [[PubMed](#)]
6. Van der Valk, T.; Pečnerová, P.; Díez-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. [[CrossRef](#)]
7. Horneck, G.; Klaus, D.M.; Mancinelli, R.L. Space Microbiology. *Microbiol. Mol. Biol. Rev.* **2010**, *74*, 121–156. [[CrossRef](#)]
8. Horneck, G.; Bücker, H.; Reitz, G. Long-Term Survival of Bacterial Spores in Space. *Adv. Space Res.* **1994**, *14*, 41–45. [[CrossRef](#)]
9. Cadet, J.; Sage, E.; Douki, T. Ultraviolet Radiation-Mediated Damage to Cellular DNA. *Mutat. Res.* **2005**, *571*, 3–17. [[CrossRef](#)]
10. Xue, Y.; Nicholson, W.L. The Two Major Spore DNA Repair Pathways, Nucleotide Excision Repair and Spore Photoproduct Lyase, Are Sufficient for the Resistance of *Bacillus Subtilis* Spores to Artificial UV-C and UV-B but Not to Solar Radiation. *Appl. Environ. Microbiol.* **1996**, *62*, 2221–2227. [[CrossRef](#)]
11. Sancho, L.G.; de la Torre, R.; Horneck, G.; Ascaso, C.; de Los Rios, A.; Pintado, A.; Wierzechos, J.; Schuster, M. Lichens Survive in Space: Results from the 2005 LICHENS Experiment. *Astrobiology* **2007**, *7*, 443–454. [[CrossRef](#)]
12. Gauslaa, Y.; Solhaug, K.A. Photoinhibition in Lichens Depends on Cortical Characteristics and Hydration. *Lichenologist* **2004**, *36*, 133–143. [[CrossRef](#)]
13. Ahmed, R.K.; Mohammed, I.J. Developing a New Hybrid Cipher Algorithm Using DNA and RC4. *Int. J. Adv. Comput. Sci. Appl.* **2017**, *8*, 71.
14. Zhang, Y.; Kong, L.; Wang, F.; Li, B.; Ma, C.; Chen, D.; Liu, K.; Fan, C.; Zhang, H. Information Stored in Nanoscale: Encoding Data in a Single DNA Strand with Base64. *Nano Today* **2020**, *33*, 100871. [[CrossRef](#)]
15. Church, G.M.; Gao, Y.; Kosuri, S. Next-Generation Digital Information Storage in DNA. *Science* **2012**, *337*, 1628. [[CrossRef](#)]
16. 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. [[CrossRef](#)]
17. Ailenberg, M.; Rotstein, O.D. An Improved Huffman Coding Method for Archiving Text, Images, and Music Characters in DNA. *BioTechniques* **2009**, *47*, 747–754. [[CrossRef](#)]
18. Yazdi, S.M.H.T.; Gabrys, R.; Milenkovic, O. Portable and Error-Free DNA-Based Data Storage. *Sci. Rep.* **2017**, *7*, 5011. [[CrossRef](#)]
19. Shipman, S.L.; Nivala, J.; Macklis, J.D.; Church, G.M. CRISPR-Cas Encoding of a Digital Movie into the Genomes of a Population of Living Bacteria. *Nature* **2017**, *547*, 345–349. [[CrossRef](#)]
20. Bornholt, J.; Lopez, R.; Carmean, D.M.; Ceze, L.; Seelig, G.; Strauss, K. A DNA-Based Archival Storage System. In Proceedings of the Twenty-First International Conference on Architectural Support for Programming Languages and Operating Systems, Atlanta, GA, USA, 2–6 April 2016; ACM: Atlanta, GA, USA, 2016; pp. 637–649.
21. Blawat, M.; Gaedke, K.; Hütter, I.; Chen, X.-M.; Turczyk, B.; Inverso, S.; Pruitt, B.W.; Church, G.M. Forward Error Correction for DNA Data Storage. *Procedia Comput. Sci.* **2016**, *80*, 1011–1022. [[CrossRef](#)]
22. 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. *Nat. Biotechnol.* **2018**, *36*, 242–248. [[CrossRef](#)]
23. Choi, Y.; Ryu, T.; Lee, A.C.; Choi, H.; Lee, H.; Park, J.; Song, S.-H.; Kim, S.; Kim, H.; Park, W.; et al. High Information Capacity DNA-based Data Storage with Augmented Encoding Characters Using Degenerate Bases. *Sci. Rep.* **2019**, *9*, 6582. [[CrossRef](#)] [[PubMed](#)]
24. Lee, H.H.; Kalhor, R.; Goela, N.; Bolot, J.; Church, G.M. Terminator-Free Template-Independent Enzymatic DNA Synthesis for Digital Information Storage. *Nat. Commun.* **2019**, *10*, 2383. [[CrossRef](#)] [[PubMed](#)]
25. Tabatabaei, S.K.; Wang, B.; Athreya, N.B.M.; Enghiad, B.; Hernandez, A.G.; Fields, C.J.; Leburton, J.-P.; Soloveichik, D.; Zhao, H.; Milenkovic, O. DNA Punch Cards for Storing Data on Native DNA Sequences via Enzymatic Nicking. *Nat. Commun.* **2020**, *11*, 1742. [[CrossRef](#)] [[PubMed](#)]
26. Yang, K.; McCloskey, C.M.; Chaput, J.C. Reading and Writing Digital Information in TNA. *ACS Synth. Biol.* **2020**, *9*, 2936–2942. [[CrossRef](#)] [[PubMed](#)]
27. Ren, Y.; Zhang, Y.; Liu, Y.; Wu, Q.; Su, J.; Wang, F.; Chen, D.; Fan, C.; Liu, K.; Zhang, H. DNA-Based Concatenated Encoding System for High-Reliability and High-Density Data Storage. *Small Methods* **2022**, *6*, e2101335. [[CrossRef](#)]
28. Mayer, C.; McNroy, G.R.; Murat, P.; Van Delft, P.; Balasubramanian, S. An Epigenetics-Inspired DNA-Based Data Storage System. *Angew. Chem. Int. Ed.* **2016**, *55*, 11144–11148. [[CrossRef](#)]
29. Sinyakov, A.N.; Ryabinin, V.A.; Kostina, E.V. Application of Array-Based Oligonucleotides for Synthesis of Genetic Designs. *Mol. Biol.* **2021**, *55*, 487–500. [[CrossRef](#)]
30. Song, L.-F.; Deng, Z.-H.; Gong, Z.-Y.; Li, L.-L.; Li, B.-Z. Large-Scale de Novo Oligonucleotide Synthesis for Whole-Genome Synthesis and Data Storage: Challenges and Opportunities. *Front. Bioeng. Biotechnol.* **2021**, *9*, 689797. [[CrossRef](#)]
31. Heckel, R.; Shomorony, I.; Ramchandran, K.; Tse, D.N.C. Fundamental Limits of DNA Storage Systems. In Proceedings of the 2017 IEEE International Symposium on Information Theory (ISIT), Aachen, Germany, 25–30 June 2017; pp. 3130–3134.
32. Zhang, Y.; Ren, Y.; Liu, Y.; Wang, F.; Zhang, H.; Liu, K. Preservation and Encryption in DNA Digital Data Storage. *Chempluschem* **2022**, *87*, e202200183. [[CrossRef](#)]
33. 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. *Nat. Protoc.* **2020**, *15*, 86–101. [[CrossRef](#)]

34. Xie, R.; Zan, X.; Chu, L.; Su, Y.; Xu, P.; Liu, W. Study of the Error Correction Capability of Multiple Sequence Alignment Algorithm(MAFFT) in DNA Storage. *BMC Bioinform.* **2023**, *24*, 111. [[CrossRef](#)]
35. Erlich, Y.; Zielinski, D. DNA Fountain Enables a Robust and Efficient Storage Architecture. *Science* **2017**, *355*, 950–954. [[CrossRef](#)]
36. Tan, X.; Ge, L.; Zhang, T.; Lu, Z. Preservation of DNA for Data Storage. *Russ. Chem. Rev.* **2021**, *90*, 280–291. [[CrossRef](#)]
37. 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. [[CrossRef](#)]
38. Paunescu, D.; Puddu, M.; Soellner, J.O.B.; Stoessel, P.R.; Grass, R.N. Reversible DNA Encapsulation in Silica to Produce ROS-Resistant and Heat-Resistant Synthetic DNA “Fossils”. *Nat. Protoc.* **2013**, *8*, 2440–2448. [[CrossRef](#)]
39. 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. *Nat. Commun.* **2019**, *10*, 1706. [[CrossRef](#)]
40. Choi, Y.; Bae, H.J.; Lee, A.C.; Choi, H.; Lee, D.; Ryu, T.; Hyun, J.; Kim, S.; Kim, H.; Song, S.-H.; et al. DNA Micro-Disks for the Management of DNA-Based Data Storage with Index and Write-Once-Read-Many(WORM) Memory Features. *Adv. Mater.* **2020**, *32*, e2001249. [[CrossRef](#)]
41. Anchordoquy, T.J.; Molina, M.C. Preservation of DNA. *Cell Preserv. Technol.* **2007**, *5*, 180–188. [[CrossRef](#)]
42. Ivanova, N.V.; Kuzmina, M.L. Protocols for Dry DNA Storage and Shipment at Room Temperature. *Mol. Ecol. Resour.* **2013**, *13*, 890–898. [[CrossRef](#)]
43. Chen, W.D.; Kohll, A.X.; Nguyen, B.H.; Koch, J.; Heckel, R.; Stark, W.J.; Ceze, L.; Strauss, K.; Grass, R.N. Combining Data Longevity with High Storage Capacity—Layer-by-Layer DNA Encapsulated in Magnetic Nanoparticles. *Adv. Funct. Mater.* **2019**, *29*, 1901672. [[CrossRef](#)]
44. Kim, T.W.; Kim, I.Y.; Park, D.-H.; Choy, J.-H.; Hwang, S.-J. Highly Stable Nanocontainer of APTES-Anchored Layered Titanate Nanosheet for Reliable Protection/Recovery of Nucleic Acid. *Sci. Rep.* **2016**, *6*, 21993. [[CrossRef](#)] [[PubMed](#)]
45. Frantzen, M.a.J.; Silk, J.B.; Ferguson, J.W.H.; Wayne, R.K.; Kohn, M.H. Empirical Evaluation of Preservation Methods for Faecal DNA. *Mol. Ecol.* **1998**, *7*, 1423–1428. [[CrossRef](#)] [[PubMed](#)]
46. Kilpatrick, C.W. Noncryogenic Preservation of Mammalian Tissues for DNA Extraction: An Assessment of Storage Methods. *Biochem. Genet.* **2002**, *40*, 53–62. [[CrossRef](#)] [[PubMed](#)]
47. Murphy, M.A.; Waits, L.P.; Kendall, K.C.; Wasser, S.K.; Higbee, J.A.; Bogden, R. An Evaluation of Long-Term Preservation Methods for Brown Bear (*Ursus Arctos*) Faecal DNA Samples. *Conserv. Genet.* **2002**, *3*, 435–440. [[CrossRef](#)]
48. Vitošević, K.; Todorović, M.; Slović, Ž.; Varljen, T.; Matić, S.; Todorović, D. DNA Isolated from Formalin-Fixed Paraffin-Embedded Healthy Tissue after 30 Years of Storage Can Be Used for Forensic Studies. *Forensic. Sci. Med. Pathol.* **2021**, *17*, 47–57. [[CrossRef](#)]
49. Ferrer, I.; Armstrong, J.; Capellari, S.; Parchi, P.; Arzberger, T.; Bell, J.; Budka, H.; Ströbel, T.; Giaccone, G.; Rossi, G.; et al. Effects of Formalin Fixation, Paraffin Embedding, and Time of Storage on DNA Preservation in Brain Tissue: A BrainNet Europe Study. *Brain Pathol.* **2007**, *17*, 297–303. [[CrossRef](#)]
50. Smith, S.; Morin, P.A. Optimal Storage Conditions for Highly Dilute DNA Samples: A Role for Trehalose as a Preserving Agent. *J. Forensic. Sci.* **2005**, *50*, 1101–1108. [[CrossRef](#)]
51. Nguyen, H.H.; Park, J.; Park, S.J.; Lee, C.-S.; Hwang, S.; Shin, Y.-B.; Ha, T.H.; Kim, M. Long-Term Stability and Integrity of Plasmid-Based DNA Data Storage. *Polymers* **2018**, *10*, 28. [[CrossRef](#)]
52. 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. *Proc. Biol. Sci.* **2012**, *279*, 4724–4733. [[CrossRef](#)]
53. Chaorattanakawee, S.; Natalang, O.; Hananantachai, H.; Nacher, M.; Brockman, A.; Krudsood, S.; Looreesuwan, S.; Patarapotikul, J. Storage Duration and Polymerase Chain Reaction Detection of Plasmodium Falciparum from Blood Spots on Filter Paper. *Am. J. Trop. Med. Hyg.* **2003**, *69*, 42–44. [[CrossRef](#)]
54. Saieg, M.A.; Geddie, W.R.; Boerner, S.L.; Liu, N.; Tsao, M.; Zhang, T.; Kamel-Reid, S.; da Cunha Santos, G. The Use of FTA Cards for Preserving Unfixed Cytological Material for High-Throughput Molecular Analysis. *Cancer Cytopathol.* **2012**, *120*, 206–214. [[CrossRef](#)]
55. 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. *Nat. Biotechnol.* **2020**, *38*, 39–43. [[CrossRef](#)]
56. Antkowiak, P.L.; Koch, J.; Rzepka, P.; Nguyen, B.H.; Strauss, K.; Stark, W.J.; Grass, R.N. Anhydrous Calcium Phosphate Crystals Stabilize DNA for Dry Storage. *Chem. Commun.* **2022**, *58*, 3174–3177. [[CrossRef](#)]
57. Coudy, D.; Colotte, M.; Luis, A.; Tuffet, S.; Bonnet, J. Long Term Conservation of DNA at Ambient Temperature. Implications for DNA Data Storage. *PLoS ONE* **2021**, *16*, e0259868. [[CrossRef](#)]
58. Clermont, D.; Santoni, S.; Saker, S.; Gomard, M.; Gardais, E.; Bizet, C. Assessment of DNA Encapsulation, a New Room-Temperature DNA Storage Method. *Biopreserv. Biobank.* **2014**, *12*, 176–183. [[CrossRef](#)]
59. Organick, L.; Nguyen, B.H.; McAmis, R.; Chen, W.D.; Kohll, A.X.; Ang, S.D.; Grass, R.N.; Ceze, L.; Strauss, K. An Empirical Comparison of Preservation Methods for Synthetic DNA Data Storage. *Small Methods* **2021**, *5*, 2001094. [[CrossRef](#)]
60. Evans, R.K.; Xu, Z.; Bohannon, K.E.; Wang, B.; Bruner, M.W.; Volkin, D.B. Evaluation of Degradation Pathways for Plasmid Dna in Pharmaceutical Formulations via Accelerated Stability Studies. *J. Pharm. Sci.* **2000**, *89*, 76–87. [[CrossRef](#)]

61. 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. [[CrossRef](#)]
62. 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. [[CrossRef](#)]
63. 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 Res.* **2010**, *38*, 1531–1546. [[CrossRef](#)]
64. Cherng, J.-Y.; Talsma, H.; Crommelin, D.J.A.; Hennink, W.E. Long Term Stability of Poly((2-Dimethylamino)Ethyl Methacrylate)-Based Gene Delivery Systems. *Pharm. Res.* **1999**, *16*, 1417–1423. [[CrossRef](#)] [[PubMed](#)]
65. Molina, M.D.C.; Anchordoquy, T.J. Degradation of Lyophilized Lipid/DNA Complexes during Storage: The Role of Lipid and Reactive Oxygen Species. *Biochim. Biophys. Acta Biomembr.* **2008**, *1778*, 2119–2126. [[CrossRef](#)] [[PubMed](#)]
66. Zhou, L.; Lei, Q.; Guo, J.; Gao, Y.; Shi, J.; Yu, H.; Yin, W.; Cao, J.; Xiao, B.; Andreo, J.; et al. Long-Term Whole Blood DNA Preservation by Cost-Efficient Cryosilicification. *Nat. Commun.* **2022**, *13*, 6265. [[CrossRef](#)] [[PubMed](#)]
67. Hao, M.; Qiao, H.; Gao, Y.; Wang, Z.; Qiao, X.; Chen, X.; Qi, H. A Mixed Culture of Bacterial Cells Enables an Economic DNA Storage on a Large Scale. *Commun. Biol.* **2020**, *3*, 416. [[CrossRef](#)]
68. Lee, H.; Popodi, E.; Tang, H.; Foster, P.L. Rate and Molecular Spectrum of Spontaneous Mutations in the Bacterium *Escherichia coli* as Determined by Whole-Genome Sequencing. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E2774–E2783. [[CrossRef](#)]
69. Chen, W.; Han, M.; Zhou, J.; Ge, Q.; Wang, P.; Zhang, X.; Zhu, S.; Song, L.; Yuan, Y. An Artificial Chromosome for Data Storage. *Natl. Sci. Rev.* **2021**, *8*, nwab028. [[CrossRef](#)]
70. Zhou, J.; Zhang, C.; Wei, R.; Han, M.; Wang, S.; Yang, K.; Zhang, L.; Chen, W.; Wen, M.; Li, C.; et al. Exogenous artificial DNA forms chromatin structure with active transcription in yeast. *Sci. China Life Sci.* **2021**, *65*, 851–860. [[CrossRef](#)]
71. Meas, R.; Wyrick, J.J.; Smerdon, M.J. Nucleosomes regulate base excision repair in chromatin. *Mutat. Res.-Rev. Mutat. Res.* **2019**, *780*, 29–36. [[CrossRef](#)]
72. Sun, Z.; Zhang, Y.; Jia, J.; Fang, Y.; Tang, Y.; Wu, H.; Fang, D. H3K36me3, message from chromatin to DNA damage repair. *Cell Biosci.* **2020**, *10*, 9. [[CrossRef](#)]
73. Hao, Y.; Li, Q.; Fan, C.; Wang, F. Data Storage Based on DNA. *Small Struct.* **2021**, *2*, 2000046. [[CrossRef](#)]
74. Liu, Y.; Ren, Y.; Li, J.; Wang, F.; Wang, F.; Ma, C.; Chen, D.; Jiang, X.; Fan, C.; Zhang, H.; et al. In Vivo Processing of Digital Information Molecularly with Targeted Specificity and Robust Reliability. *Sci. Adv.* **2022**, *8*, eabo7415. [[CrossRef](#)]
75. Al Mamun, A.A.M.; Lombardo, M.-J.; Shee, C.; Lisewski, A.M.; Gonzalez, C.; Lin, D.; Nehring, R.B.; Saint-Ruf, C.; Gibson, J.L.; Frisch, R.L.; et al. Identity and function of a large gene network underlying mutagenic repair of DNA breaks. *Science* **2012**, *338*, 1344–1348. [[CrossRef](#)]
76. Oller, A.R.; Schaaper, R.M. Spontaneous mutation in *Escherichia coli* containing the dnaE911 DNA polymerase antimutator allele. *Genetics* **1994**, *138*, 263–270. [[CrossRef](#)]
77. Schaaper, R.M. Suppressors of *Escherichia coli* mutT: Antimutators for DNA replication errors. *Mutat. Res.* **1996**, *350*, 17–23. [[CrossRef](#)]
78. Woo, A.C.; Faure, L.; Dapa, T.; Matic, I. Heterogeneity of spontaneous DNA replication errors in single isogenic *Escherichia coli* cells. *Sci. Adv.* **2018**, *4*, eaat1608. [[CrossRef](#)]
79. Tabatabaei, S.K.; Pham, B.; Pan, C.; Liu, J.; Chandak, S.; Shorkey, S.A.; Hernandez, A.G.; Aksimentiev, A.; Chen, M.; Schroeder, C.M.; et al. Expanding the Molecular Alphabet of DNA-Based Data Storage Systems with Neural Network Nanopore Readout Processing. *Nano Lett.* **2022**, *22*, 1905–1914. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.