VIVA Institute of Technology
9thNational Conference onRole of Engineers in Nation Building – 2021 (NCRENB-2021)

VIVA-TECH IJRI
ANNUAL RESEARCH JOURNAL

VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION
ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

An Alternative to Hard Drives in the Coming Future: DNA-BASED DATA STORAGE

Rajeshri Deshmukh[1], Chandani Patel[2]
[1]*Department of Computer Applications, VIVA School of MCA, India*
[2]*Department of Computer Applications, VIVA School of MCA, India*

***Abstract*--** *Demand for data storage is growing exponentially, but the capacity of existing storage media is not keeping up, there emerges a requirement for a storage medium with high capacity, high storage density, and possibility to face up to extreme environmental conditions. According to a research in 2018, every minute Google conducted 3.88 million searches, other people posted 49,000 photos on Instagram, sent 159,362,760 e-mails, tweeted 473,000 times and watched 4.33 million videos on YouTube. In 2020 it estimated a creation of 1.7 megabytes of knowledge per second per person globally, which translates to about 418 zettabytes during a single year. The magnetic or optical data-storage systems that currently hold this volume of 0s and 1s typically cannot last for quite a century. Running data centres takes vast amounts of energy. In short, we are close to have a substantial data-storage problem which will only become more severe over time. Deoxyribonucleic acid (DNA) are often potentially used for these purposes because it isn't much different from the traditional method utilized in a computer. DNA's information density is notable, 215 petabytes or 215 million gigabytes of data can be stored in just one gram of DNA. First we can encode all data at a molecular level and then store it in a medium that will last for a while and not become out-dated just like floppy disks. Due to the improved techniques for reading and writing DNA, a rapid increase is observed in the amount of possible data storage in DNA.*

Keywords: data storage, DNA, floppy disk, information density, optical data storage systems

I. INTRODUCTION

The outing of information stockpiling started from bones, shakes, and paper. At that point this excursion digressed to punched cards, attractive tapes, gramophone records, and floppies, at that point forward. Thereafter with the occasion of the innovation optical circles including CDs, DVDs, Blue-beam plates, and blaze drives came into activity. Those are exposed to rot. Being non-biodegradable materials these contaminate the climate and furthermore discharge high measures of warmth energy while utilizing energy for activity.

Consistently in 2018, Google led 3.88 million pursuits, and people watched 4.33 million recordings on YouTube, sent 159,362,760 messages, tweeted multiple times and posted 49,000 photographs on Instagram, in sync with programming organization Domo. By 2020 an expected 1.7 megabytes of data are visiting be made every second per individual worldwide, which means around 418 zettabytes during one year (418 billion one-terabyte hard drive of data), accepting a total populace of seven.8 billion[3].

The attractive or optical information stockpiling frameworks that presently hold this volume of 0s and 1s normally can't keep going for longer than a century, if that. Further, running server farms takes gigantic measures of energy. To put it plainly, we are on the purpose of have a weighty information stockpiling issue that may just turn out to be more extreme over the long haul. Interest for information stockpiling is developing

dramatically, however the limit of existing stockpiling media isn't keeping up. A large portion of the world's information today is put away on attractive and optical media.

Regardless of enhancements in optical plates, putting away a zettabyte of information would in any case take a large number of units, and utilize huge actual space. In the event that we are to safeguard the world's information, we need to look for critical advances away thickness and sturdiness[3]. Utilizing DNA to document information is an alluring chance since it is amazingly thick (up to around 1 exabyte for each cubic millimetre) and strong (half-existence of more than 500 years).

II. LITERATURE REVIEW

DNA, which consists of long nucleotide chains A, T, C and G, is that the information-storage material of life. Within the sequence of those letters, data are often processed, turning DNA into a replacement sort of information technology. It's already sequenced (read) daily, synthesized (written to) and quickly replicated accurately. As has been shown by the complete genome sequencing of a fossil horse that lived quite 500,000 years ago, DNA is additionally remarkably stable. And it doesn't take much energy to store it. But it's the potential for storage that shines. At a density far exceeding that of electronic devices, DNA can accurately store vast quantities of knowledge.For example, consistent with calculations published in 2016 in Nature Materials by George Church of Harvard University and his colleagues, the straightforward bacterium Escherichia coli features a storage density of about 1019 bits per millilitre. At that density, a cube of DNA measuring around one meter on one side might well fulfil all the world's present storage needs for a year.

2.1 REVIEW OF PREVIOUS STUDIES

PCs and other computerized electronic gadgets store information and work with the double numeric framework that utilizes just two advanced numbers or 0 and 1. The writings are changed over to parallel variant in PC framework. Thus, PCs work, and compute in twofold, in the end convert data to messages lucid. One byte contains eight pieces comprising of one or the other 0's or 1's and having 28 (256) potential qualities (from 0 to 255), and stores one single letter (Figure 1 and Table 1). As demonstrated in the transformation ASCII (Table 1). The 26 letters with the upper cases and lower cases are changed over among Letter, Binary and Hexadecimal. To store an enormous record or archive need substantially more memory information. An ordinary tune may require many megabytes, with couple gigabytes to store a film and a few terabytes for the books put away in a huge library[2].

As demonstrated in Table 2 are the extents of estimation and memory for the utilization of paired framework from the littlest unit "byte" to the huge units, including byte (B), kilobyte (KB), megabyte (MB), gigabyte (GB), terabyte (TB), pegabyte (PB), Exabyte (EB), zettabyte (ZB), yottabyte (YB), brontobyte (BB), Geopbyte (GPB, etc. The units like brontobyte (BB), Geopbyte (GPB) are unfathomable colossal qualities that may never be utilized in our genuine world (Table 2).

TABLE 1

The conversion ascii table of the twenty-six letters with the upper and lower cases among letter, binary and hexadecimal.

Letter	Binary	Hexadecimal	Letter	Binary	Hexadecimal
A	1000001	41	a	1100001	61
B	1000010	42	b	1100010	62
C	1000011	43	c	1100011	63
D	1000100	44	d	1100100	64
E	1000101	45	e	1100101	65
F	1000110	46	f	1100110	66
G	1000111	47	g	1100111	67
H	1001000	48	h	1101000	68
I	1001001	49	i	1101001	69
J	1001010	4A	j	1101010	6A
K	1001011	4B	k	1101011	6B
L	1001100	4C	l	1101100	6C

Letter	Binary	Hexadecimal	Letter	Binary	Hexadecimal
M	1001101	4D	m	1101101	6D
N	1001110	4E	n	1101110	6E
O	1001111	4F	o	1101111	6F
P	1010000	50	p	1110000	70
Q	1010001	51	q	1110001	71
R	1010010	52	r	1110010	72
S	1010011	53	s	1110011	73
T	1010100	54	t	1110100	74
U	1010101	55	u	1110101	75
V	1010110	56	v	1110110	76
W	1010111	57	w	1110111	77
X	1011000	58	x	1111000	78
Y	1011001	59	y	1111001	79
Z	1011010	5A	z	1111010	7A

TABLE 2
The sizes of measurement and memory

Sizes	Byte Magnitude	Units	Storage*
1 B	100	Byte	A character “A”, “1”, “$”
10 B	101		
100 B	102		
1 KB	103	Kilo byte	The size for graphics of small websites ranges between 5 and 100 KB
10 KB	104		
100 KB	105		
1 MB	106	Mega byte (1 MB: 1 million)	The size for a high resolution JPEG image is about 1-5 MB
10 MB	107		The size for a 3-minute song is about 30 MB
100 MB	108		
1 GB	109	Giga byte	The size for a standard DVD drive is about 5 GB
10 GB	1010	(1 GB: 1 billion)	
100 GB	1011		
1 TB	1012	Tera byte (1 TB: 1 trillion)	The size for a typical internal HDD is about 2 TB
10 TB	1013		
100 TB	1014		

2.2 PURPOSE

With the remarkable development in the limit of data produced and the arising need for information to be put away for delayed timeframe, there arises a requirement for a capacity medium with high limit, high stockpiling thickness, and plausibility to withstand extraordinary ecological conditions.

DNA arises because the approaching automobile for facts stockpiling with its putting highlights. DNA has a extraordinary stockpiling limit. Castillo states that everyone the information withinside the entire Internet may be located in a system that is lesser than unit cubic inch. DNA is visible as the right medium in such way in a fashionable feel in mild of the reality that rather than making use of 1 s and 0 s through the PC to save facts, DNA comprising of adenine, guanine, cytosine, and thymine (A, G, C, and T) successfully matched into nucleotide base combines A-T and G-C may be used for placing away information in a form of fold code.
DNA arises as the planned mechanism for information stockpiling with its striking highlights. DNA has an incredible stockpiling limit. Castillo states that all the data in the whole Internet could be situated in a gadget which is lesser than unit cubic inch. DNA is seen as the ideal medium in such manner generally in light of the

fact that as opposed to utilizing 1 s and 0 s by the PC to store information, DNA comprising of adenine, guanine, cytosine, and thymine (A, G, C, and T) effectively combined into nucleotide base matches A-T and G-C can be used for putting away data in a type of twofold code.

As the urgent need for top capacity data-storage medium rises, DNA is taken into account ideal during this think of single nucleotide can represent 2 bits of data. Accordingly 455 EB of knowledge are often encoded in 1 gram of single stranded DNA (ssDNA). Entire information that's produced by the planet over a year are often stored in only 4 grams of DNA.

High memory space is obtainable by DNA because it is 3-dimensional (3D) by structure. DNA offers readable and reliable information for millennia, which may be extended to almost infinity by drying and protecting from oxygen and water [3]. DNA can withstand a broader range of temperatures (−800°C–800°C). It utilizes power usage million times more effectively than a contemporary pc. Additionally it privileges more storage options because it stores data during a nonlinear structure unlike most of the media storing data during a linear structure. DNA promises more options to enhance latency and extraction of knowledge, because it allows reading data in bi-directions. The important incontrovertible fact that DNA is invisible to human eye ensures that DNA is secure and is impossible to be harmed by living organisms[3].

III. WORKING OF DNA STORAGE

Lately computerized information is a vital piece of our life. Our own information, for example, individual data, our computerized keys, advanced wallet data, passwords and bank subtleties are a portion of the vital information that should be put away safely.

The advanced information is encoded in a DNA arrangement, the relating grouping data is incorporated into a fake DNA and the data is decoded by sequencing the counterfeit DNA strand. This is the specific way of putting away and recovering computerized information from DNA.

3.1 ENCODING DATA INTO THE DNA SEQUENCE

The computer is laboured on a binary gadget of one and 2. In the first actual step, virtual information is included into the DNA. The DNA has four nitrogenous bases: Adenine (A), Cytosine (C), Guanine (G) and Thymine (T). For storing information into the DNA, the A, T, G and C bases of DNA first transformed into binary codes 1 and 0. 00 for A, 01 for G, 10 for C and eleven for T are the binary codes for storing facts. The facts withinside the binary shape is transformed into the collection of A, T, G, C. Now we've the lengthy virtual collection of DNA.

3.1.1 CODES FOR ENCRYPTING DATA IN DNA

In the past, basically, 3 codes were used to store DNA information. In general, both of these codes consider that an alphabetic language is encoded in DNA[4]. Since most of the studies considered English as the alphabetic language, the writing scheme for phonetics may have been used for shorthand.

It should meet the dual requirements as follows in order for a code to be optimum:

1) DNA (nucleotides) should be used commercially, primarily because synthesizing extended oligonucleotides is an expensive operation, although it appears to be reasonably economical to replicate.
2) After data encoding, it should be able to reconstruct the message.

Although it's not considered to be essential, if the coding scheme offers some error detection and protection mechanism it might be of tremendous advantage. But this feature isn't considered vitally important, because there are other mechanisms for addressing this issue like using multiple copies of DNA. Because the written communication inherently consists of self-correcting mechanisms it makes this feature of error detection and correction not essentially important [4].

Huffman Coding -This code uses the principle of varying the length of symbols used for representing a personality. Most recurrently appearing character within the text is assigned rock bottom number of symbols while the smallest amount recurrently appearing character is assigned the foremost number of symbols. Employing this principle results in developing of a really economical code. Average code length is around 2.2 characters in Huffman coding scheme[8]. This is often the smallest amount average codon length achieved.

Unambiguity of the code is achieved through comprising of just one way during which the encrypted message are often read once the start line is mentioned.

Disadvantages related with Huffman coding incorporate not equipping for numbers and images. This is mostly in light of the fact that the recurrence of demonstrating these images is exceptionally reliant on the content which surveys the way that they can't be remembered for forming the Huffman code. Besides it isn't reasonable for long haul stockpiling because of the way that when distinctive length codons amassed together it probably won't uncover an example. Hence the people in the future probably won't have the option to identify the meaning of the example[8].

The Comma Code - In this methodology a solitary base (G) is considered as the comma. Codons of 5-base length are isolated from one another utilizing base G. % base codons comprise of other three bases, specifically, A, C, and T, further more restricted to single A:T base pair and two G:C base sets. The C of the second G-C pair is constantly situated in the upper strand.

Consisting of isothermal melting temperature is the benefit of the composition of the message DNA making use of this scheme. Dominant characteristic of comma code is the studying body of six codons such as G, the comma, which isn't always executed through different codes. This enables to discover a clean studying body without the need to say a beginning point. Protection mechanism from insertion and deletion mutation is likewise assured through this technique which makes the alternative codes plenty greater complex.

Drawback of this code is that it's not economical because it repeats the comma-base G to make an automatic reading frame. The Alternating Code. This scheme consists of 6 base codons which are 64 in number including pyrimidine and purines. Construction of the message DNA in a completely synthetic nature is that the primary feature of this approach. As this creates fully artificial DNA it's suitable for future storage which overcomes the disadvantage of Huffman code. Additionally, it offers benefits like being isothermal and error detecting but it's not superior to comma code.

Alternating code also comprises repetitive features which makes it non-economical. It's the most drawback related to this coding scheme. Therefore attention of the researchers has been led towards developing a cheap code without repetitive features.
Comma-Free Code - It's also referred to as prefix free code. This comprises fixed length base frames without commas to separate the frames. Therefore, it uses an automatic frame detection mechanism. Comma-free code doesn't contains identical four base pairs which is that the only way of hindering from natural DNA sequences.
These codons are possible to be read simply in a method and support error detection mechanisms also. Although comma-free code is strong and therefore the error correction works to correct against small-scale loss like DNA point mutations, it doesn't have the power to recover broken data when an outsized DNA segment is deleted from the info encoded DNA region.

3.2 ARTIFICIAL DNA SYNTHESIS

The single-abandoned self-assertive DNA grouping can be combined synthetically. Based on the computerized succession information, every nucleotide is added to the adjoining nucleotide. Notwithstanding, the proficiency of counterfeit DNA blend is 99% however the blunder of 1% can make a significant issue in advanced information stockpiling. To conquer this issue, huge quantities of equal beginning locales are given to deliver various duplicates of the given succession. In this manner, in spite of having a blunder in a solitary duplicate numerous other precise can be delivered.

3.3 STORING OF SAMPLE

Presently we have our information reinforcement as a fluid drop of a few nano-grams of DNA. The DNA can be put away in profound freeze where it tends to be keep going for a very long time or we can send it to the outer stockpiling frameworks (given by certain organizations) which can store our DNA for over thousand years.
DNA stays stable in any brutal conditions for a long period of time. Regardless, a few successions could be lost throughout some stretch of time.

3.4 SEQUENCING OF DNA

For extracting the virtual facts again to its authentic form, we must collection the complete DNA. DNA sequencing is a manner wherein a DNA collection is study into the virtual collection. The labelled nucleotides

are brought complementary to our DNA strand. Each nucleotide is labelled with a different fluorescent dye. The depth of color emitted through every dye is recorded through the detector.

3.5 DECODING INFORMATION

Finally, the sequence gets back to the decoder which decodes the DNA sequence back into binary language. After decoding, we can retrieve our data back.

TABLE 3
Comparison of data storage units with respect to access time and durability.

Durability	Data storage unit	Access time
3 years	Flash drive	Mili second
5 years	HDD (hard disk)	10 second
Up to 30 years	Magnetic tape	1 minute
More than 100 years	DNA storage	More than 12 hours

IV. ADVANTAGES OF DNA DATA STORAGE MEDIUM

The worldwide data square measure powerfully dilated at the dramatic rate. The customary media cannot adequately manage the need of the big data storage. DNA could fill in as a possible mechanism of processed data storage, with its potential favourable circumstances, for instance, high thickness, high replication effectiveness, end of the day strength and end of the day solidness [9].

DNA at its theoretic greatest limit will encrypt around 2 items for each ester. An entire server farm worked by IBM in 2011 has around a hundred petabytes (PBs) of data golf stroke away limit. All the same, because of having a high thickness, DNA going regarding as associate data golf stroke away medium will store plenty of data at a touch size [11].

A solitary gram of DNA at its theoretic most extreme will store around two hundred PBs of data, much twofold occasions quite that of the full IBM server farm. All in all, all information recorded all over on the globe will be place away in an exceedingly few kilograms of DNAs, or cherish simply one shoebox contrasted and therefore the necessity of scores of huge data storage habitats for typical media. Information encoded DNA medium is ready to try to end of the day storage owing to having high strength.

DNA will keep going for millennia exposed, dry and boring spots. Abundant below additional unfortunate climate, DNA's half-life is as long as hundred years. DNA will hold stable at vasoconstrictor or heat, with the wide reach from - 800°C to 800°C [11]. DNA media will likewise confirm regarding data quite customary processed data media. Albeit new data square measure increasing at a dramatic rate, the bulk of them square measure saved in files for end of the day storage.

These chilly data will not be recovered promptly or used typically. Consequently, to store them in DNA media is simple, useful and unpaid. Another little bit of leeway is that DNA is deeply saved. The traditional DNAs will exactly duplicate themselves at a high productivity and systematically with the base-matching principle (A with T, C with G). Afterward, DNA medium will deeply save data constancy for quite whereas.

V.CHALLENGES FOR DNA DATA STORAGE MEDIUM

In lightweight of its special qualities and contrasted and therefore the customary media, polymer may be the potential and promising mechanism for advanced info storage. However, it's so far to travel before polymer may be economically applied [13]. The difficulties we'd like to manage exist in several viewpoints, together with vital expense, low outturn, the restricted admittance to info storage, short factory-made oligo DNA sections, mistake rate in mix and sequencing [15].

The utilization of polymer in info storage is considerably additional pricey than the opposite standard media like tape, circle, and HDD (hard plate drive). As of now, to write and interpret info price nearly $15,000 per computer memory unit (MB). Then, current innovation in polymer uniting is restricted, with simply short oligo DNA groupings to be homogenised. The best length of every oligo DNA half is restricted to many hundred nucleotides.

During this manner, to store a solitary chronicled record, particularly one monumental document might need innumerable oligo DNAs. And moreover, the time has come back burning-through for info to compose into and get over oligo DNAs, with the contribution of assorted advances together with ever-changing info over

to parallel, secret writing twofold to oligo DNA, integration and golf stroke away polymer successions, and ill extraordinary groupings from polymer storage library, sequencing and disentangling, and finally ever-changing flex to info clear. The traditional media, for instance, circle and tape have their coherent tending to knowledge, however, oligo DNAs haven't.

Consequently, it's laborious to deal with the extraordinary encoded polymer succession that we have a tendency to hope to own. Within the interim, capricious admittance to DNA-based info storage is important, be that because it might, oligo DNAs do not have irregular access capability. Through current methodologies, simply mass access is accessible for polymer info storage. The complete DNA-based info storage ought to be organized, sequenced and decoded from polymer info storage despite the actual fact that we have a tendency to merely have to be compelled to examine a solitary computer memory unit. On these lines, the proper preliminary wont to specifically recover the proper polymer grouping is needed [14].

This can likewise provide capricious access throughout DNA sequencing and data ill. The sequencing with the novel preliminary will specifically examine simply the specified oligo DNA, rather than the complete polymer library. Also, at present, polymer combination and sequencing don't seem to be wholly impressive. Throughout polymer mix and sequencing, the event of addition, cancellation, replacement and completely different blunders is happened, with miscalculation rate being regarding I Chronicles per ester. The innovation and therefore the expense of polymer combination and sequencing don't seem to be applicable for current info storage.

VI. CONCLUSION

Despite the fact that DNA computerized information stockpiling innovation is exorbitant and tedious as of now. In any case, it will end up being extremely valuable soon. Indisputably, DNA advanced information stockpiling will be the lone expect putting away information sooner rather than later. It will alter the computerized innovation without a doubt. The ascent of DNA information stockpiling, beforehand the stuff of sci-fi, is being made conceivable by progresses in biotechnology, especially upgrades in high-throughput DNA sequencing and union.

Likewise, on the grounds that these bio-developers control what materials enter their investigations, and their arrangements don't should be carefully designed to work inside a living creature, there are less overhead expenses contrasted with average life science tests [12]. The excursion has not been without detours, be that as it may. Notwithstanding emotional improvement, working with DNA can be moderate and costly. Further smoothing out is as yet required.

This review critically analyzes the prevailing methods of storing data onto DNA. Data is encrypted into DNA using diverse codes and this text analyzes and discusses the codes used for encrypting data. Multiple approaches for designing DNA codons and diverse data storage styles are analyzed intimately identifying the pros and cons of every approach. Secret writing techniques using DNA molecules for secure data storage also are discussed through this text. DNA are often used as an organic memory to store massive amounts of knowledge. This paper also analyzes the mechanism where living organism might be used as storage devices while identifying limitations and appropriate applicability [9]. Challenges faced through trying to use organic memory concepts also are discussed through this paper. Big data storage and analytics and therefore the way it's led to DNA computing to unravel hard computational problems also are discussed here.

ACKNOWLEDGEMENTS

I am thankful to my college for giving us opportunity to make this project a success. I give my special thanks and sincere gratitude towards Prof.Chandani Patel for encouraging me to complete this research paper, guiding me and helping me through all the obstacles in the research.

Without her assistance, my research paper would have been impossible. Also I present my obligation towards all our past years teachers who have bestowed deep understanding and knowledge in us, over the past years. We are obliged to our parents and family members who always supported me greatly and encouraged me in each and every step.

REFERENCES

[1] S. Shrivastava and R. Badlani, “Data storage in DNA,” International Journal of Electrical Energy, vol. 2, no. 2, pp. 119–124, 2014.

[2] DNA and the Digital Data Storage, https://www.hsj.gr/medicine/dna-and-the-digital-data-storage.php?aid=24516

[3] How DNA data storage works, https://geneticeducation.co.in/dna-digital-data-storage/#:~:text=The%20digital%20data%20is%20encoded,retrieving%20digital%20data%20from%20DNA.

[4] DNA Data Storage Is Closer, https://www.scientificamerican.com/article/dna-data-storage-is-closer-than-you-think/

[5] G. C. Smith, C. C. Fiddes, J. P. Hawkins, and J. P. L. Cox, "Some possible codes for encrypting data in DNA," *Biotechnology Letters*, vol. 25, no. 14, pp. 1125–1130, 2003.

[6] M. K. Rogers and K. C. Seigfried-Spellar, "Digital forensics and cyber crime," in *Proceedings of the 4th International ICST Conference on Digital Forensics & Cyber Crime (ICDF2C '12)*, Lafayette, Ind, USA, October 2012.

[7] N. Yatchie, Y. Ohashi, and M. Tomita, "Stabilizing synthetic data in the DNA of living organisms," *Systems and Synthetic Biology*, vol. 2, no. 1-2, pp. 19–25, 2008.

[8] M. Ailenberg and O. D. Rotstein, "An improved Huffman coding method for archiving text, images, and music characters in DNA," *BioTechniques*, vol. 47, no. 3, pp. 747–754, 2009.

[9] Appuswamy RLK, Barbry P, Antonini M, Madderson O, Freemont P (2019) Archive: Using DNA in the DBMS storage hierarchy. CIDR 2019, Biennal Conference on Innovative Data Systems Research, California, USA.

[10] De Silva PY, GU Ganegoda (2016) New trends of digital data storage in DNA. Biomed Res Int pp: 8072463-8072472.

[11] O' Driscoll A, Sleator RD (2013) Synthetic DNA: the next generation of big data storage. Bioengineered 4: 123-1235.

[12] Bornholt J, Lopez R, Carmean DM, Ceze L, Seelig G, et al. (2016) A DNA-based archival storage system. ASPLOS 201 (21st ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Atlanta, GA).

[13] Organick L, Ang SD, Chen YJ, Lopez R, Yekhanin S, et al. (2018) Random access in large-scale DNA data storage. Nat Biotechnol 36: 242-248.

[14] Yazdi SM, Yuan Y, Ma J, Zhao H, Milenkovic O (2015) A rewritable, random-access DNA-based storage system. Sci Rep 5: 14138-14140.

[15] Ahn T, Ban H, Park H (2018) Storing digital information in the long read DNA. Genomics Inform 16: e30-35.

[16] A. J. Doig, "Improving the efficiency of the genetic code by varying the codon length—the perfect genetic code," *Journal of Theoretical Biology*, vol. 188, no. 3, pp. 355–360, 1997.