Int.j.pathol.2023;21(2):40 – 41

EDITORIAL

Storage of Data in DNA

Anwar ul Haque

Ours is super communication age. All over the world enormous amount of data is being generated every second from every phone, every computer, every device in every office, home, school, college, university, hospital, shop, industry, institution and moving vehicle etc. This gigantic and titanic volume of information has almost depleted our data storing capacity on internet. Scientists are therefore trying very hard to find out alternate and more spacious capacity.

Luckily, we have found the solution in DNA (or RNA) that is essential element of nuclei in all living earthly creatures! DNA, as everyone knows has many exceptional characteristics apart from it being ubiquitous, its language is amazingly simple with only 4 alphabets and its words contain only 3 alphabets. It has distinctively accurate replication system; it can also be artificially synthesized and can be inserted in living cells, its amazingly tough and hardy can survive thousands of years!

Perhaps for these reasons among many other known and unknown; Allah has made the wisest and the most efficient system of storing and retrieving information data in every cell of every earthly creature, may that be virus, bacteria, fungus, algae, parasite, insect, fish, bird, animal, and plant! Life is dependent on such super-efficient mechanism of storing, retrieving and flawless utilization of this information.

Beginning mid twentieth century, the knowledge of genetics has been enormously and rapidly growing leap and bound. Not only our understanding of DNA, RNA, and chromosomes but also our ability to manipulate these structures, their testing and their artificial synthesizing have been tremendously increased.

Therefore, it was almost natural to have an idea of utilizing this enormous capacity to store our own generated data in natural as well as in synthetic DNA. Eric Baum, in 1995, floated the idea that huge amount of data can be stored in a tiny amount of DNA due to its ultra-high density.[1] This led to DNA computing into the dominion of memory technology.

Soon, several methods for encoding data in DNA were explored which will optimally store the data! As DNA is very hardy structure, data can be easily kept safely for thousands of years.

DNA digital data storage is made possible through encoding and decoding binary data to and from synthesized strands of DNA; for example by 2019, all 16 GB of text from Wikipedia's English version had been encoded into synthetic DNA. [2-4] In 2021, a custom DNA data writer had been developed that was capable of writing data into DNA at 18 Mbps.

For encoding information into DNA several relatively simple methods are available which basically involve translating each letter into a corresponding "codon", consisting of a unique small sequence of nucleotides in a "lookup tables" (Reference tables).[5,6] To encode arbitrary data in DNA, the data is first converted into ternary (base 3) data rather than binary (base 2) data. Each digit (or "trit") is then converted into a nucleotide using a lookup table.[7]

The genetic code within cell can also be used to store information by engineering cells with "molecular recorders" to allow the storage and retrieval of information stored in the cell's genetic material.[8] In 2007, a device was created at the University of Arizona using addressing molecules to encode mismatch sites within a DNA strand. These mismatches were then able to be read out by performing a restriction digest, thereby recovering the data.[9]

Eurecom and Imperial College in January 2019, demonstrated the ability to store structured data in synthetic DNA. They showed how to encode structured and relational data in synthetic DNA and how to perform data processing on the DNA.[10] 5.5 petabits data can be stored in each cubic millimeter of DNA. DNA thus can also be another type of storage medium such as hard drives and magnetic tapes.

In 2013, researchers from the European Bioinformatics Institute (EBI) described the storage, retrieval, and reproduction of over five million bits of data. All the DNA files reproduced the information between 99.99% and 100% accuracy. The main innovations in this research were the use of an error-correcting encoding scheme to ensure the extremely low data-loss rate, as well as the idea of encoding the data in a series of overlapping

short oligonucleotides identifiable through a sequence-based indexing scheme. Also, the sequences of the individual strands of DNA overlapped in such a way that each region of data was repeated four times to avoid errors. Two of these four strands were constructed backwards, also with the goal of eliminating errors.

ETH Zurich highlighted the long stability of DNA encoded data in 2015. [11]

Yaniv Erlich and Dina Zielinski of Columbia University and the New York Genome Center in 2017 proposed "DNA Fountain" method to store data at a density of 215 petabytes per gram of DNA. *Shankland S reported packing of all 16GB of Wikipedia will last much longer than any current technology*. Robert F Service reported that entire world' data can be stored in one room.[12]

We had hypothesized for the first time that our spoken, writing and listening activity is primarily based on DNA of brain neurons rather than in synapses. We believe that as we read, write, or speak, the action is carried out at DNA in our neurons in continuous super hyperactivity at neuronal DNA. This seems also relates with permanent or transient storage of data at neuronal DNA that is known as memory.In Alzheimer disease loss or defective memory most likely primarily results from damage to the neurons.[13]

Obviously, these genes are kept silent in other cells to which some regard them as waste or garbage in rather arrogant manner. However, Quran clearly says that on the Day of Judgment our organs e.g., skin will speak "And they say unto their skins: Why testify ye against us? They say: Allah hath given us speech Who giveth speech to all things, and Who created you at the first, and unto Whom ye are returned."[14]

At present the DNA storage procedure is quite costly estimated per megabyte $12,400 to encode data and $220 for retrieval. However, it is hoped that with the exponential decrease in DNA synthesis and sequencing costs, as well emerging new technologies if it continues, should make the technology cost-effective for the long-term data storage in near future.[15]

References

1. Baum, E. B. (1995-04-28). "Building an associative memory vastly larger than the brain". Science. 268 (5210): 583–585.
2. Church, G. M.; Gao, Y.; Kosuri, S. (2012-08-16). "Next-Generation Digital Information Storage in DNA". Science. 337 (6102): 1628. Bibcode:2012Sci...337.1628C. doi:10.1126/science.1226355. ISSN 0036-8075. PMID 22903519. S2CID 934617.
3. Erlich, Yaniv; Zielinski, Dina (2017-03-02). "DNA Fountain enables a robust and efficient storage architecture". Science. 355 (6328): 950–54 Bibcode:2017Sci...355.950E. doi:10.1126/science.aaj2038. ISSN 0036-8075. PMID 28254941. S2CID 13470340.
4. Shankland S (29 June 2019). "Startup packs all 16GB of Wikipedia onto DNA strands to demonstrate new storage tech - Biological molecules will last a lot longer than the latest computer storage technology, Catalog believes". CNET. Retrieved 7 August 2019.
5. Roquet, Nathaniel; Bhatia, Swapnil P.; Flickinger, Sarah A.; Mihm, Sean; Norsworthy, Michael W.; Leake, Devin; Park, Hyunjun (2021-04-20). "DNA-based data storage via combinatorial assembly": 2021.04.20.440194. doi:10.1101/2021.04.20.440194. S2CID 233415483.
6. Organick, Lee; Ang, Siena Dumas; Chen, Yuan-Jyue; Lopez, Randolph; Yekhanin, Sergey; Makarychev, Konstantin; Racz, Miklos Z.; Kamath, Govinda; Gopalan, Parikshit; Nguyen, Bichlien; Takahashi, Christopher N. (March 2018). "Random access in large-scale DNA data storage". Nature Biotechnology. 36 (3): 242–248. doi:10.1038/nbt.4079. ISSN 1546-1696. PMID 29457795. S2CID 205285821.
7. Smith, Geoff C.; Fiddes, Ceridwyn C.; Hawkins, Jonathan P.; Cox, Jonathan P.L. (2003). "Some khpossible codes for encrypting data in DNA". Biotechnology Letters. 25 (14): 1125–1130. doi:10.1023/a:1024539608706. PMID 12966998. S2CID 20617098.
8. Skinner GM, Visscher K, Mansuripur M (2007-06-01). "Biocompatible Writing of Data into DNA". Journal of Bionanoscience. 1 (1): 17–21. arXiv:1708.08027. doi:10.1166/jbns.2007.005. S2CID 11241232.
9. O'Neill, S., I plant memories in seeds, New Scientist, Volume 229, Issue 3056, 16 January 2016, doi:10.1016/S0262-4079(16)30130-0
10. Blawat M, Gaedke K, Huetter I, Chen XM, Turczyk B, Inverso S, Pruitt BW, Church GM (2016). "Forward Error Correction for DNA Data Storage". Procedia Computer Science. 80: 1011–1022. doi:10.1016/j.procs.2016.05.398. = q`. `
11. "DNA could store all of the world's data in one room". Science Magazine. 2 March 2017. Retrieved 3 March 2017.
12. ServiceRFhttps://www.science.org/content/article/dna-could-store-all-worlds-data-one-room
13. Haque AU. int.j.pathol.2020;18(1) 1-2
14. Quran 41:21
15. Panjiyar JK https://medium.com/zerone-magazine/from-punch-card-to-dna-data-storage-5c15dcc4803e