AgriCos e-Newsletter
ISSN: 2582-7049
AgriCos e-Newsletter
Open Access Multidisciplinary Monthly Online Magazine
Volume: 04 Issue: 04 April 2023 Article No: 16

DNA: A Potential Tool for Data Storage

Ashwini M. N.
Ph. D. Research Scholar, Department of Biotechnology, Junagadh Agricultural University, Junagadh, Gujarat

SUMMARY

Data generated by the growing population is increasing day by day with the requirement of enormous storage space. Currently used storage facilities, with certain limitations needs an effective alternative as a solution to combat the pitfalls associated them. DNA stands as an sustainable alternative with it's ultra-compactness, hyper density, ultra-persistence and massive redundancy which have many advantages over traditional storage technologies.

INTRODUCTION

As the global population has skyrocketed over the last few decades, there is a huge demand for the storage of data generated by the same. It is estimated that the amount of data generated may reach 5000 zettabytes (~10^{24}) by 2040 (Ezekannagha *et al*. 2022). Currently data being stored in magnetic, optical or mechanical media that records and preserves digital information using computers or other devices, which probably become unrecoverable within a century or less. Data storage centres will consume large physical space, cost, electricity, requires significant cooling system and emit large amount of carbon dioxide. The retention time of most of the storage media is also short (Limbachiya et al., 2022, Ezekannagha *et al*. 2022). Consequently, there is a need for better storage media with improved information density, durability and energy costs. The solution lies in nature itself.

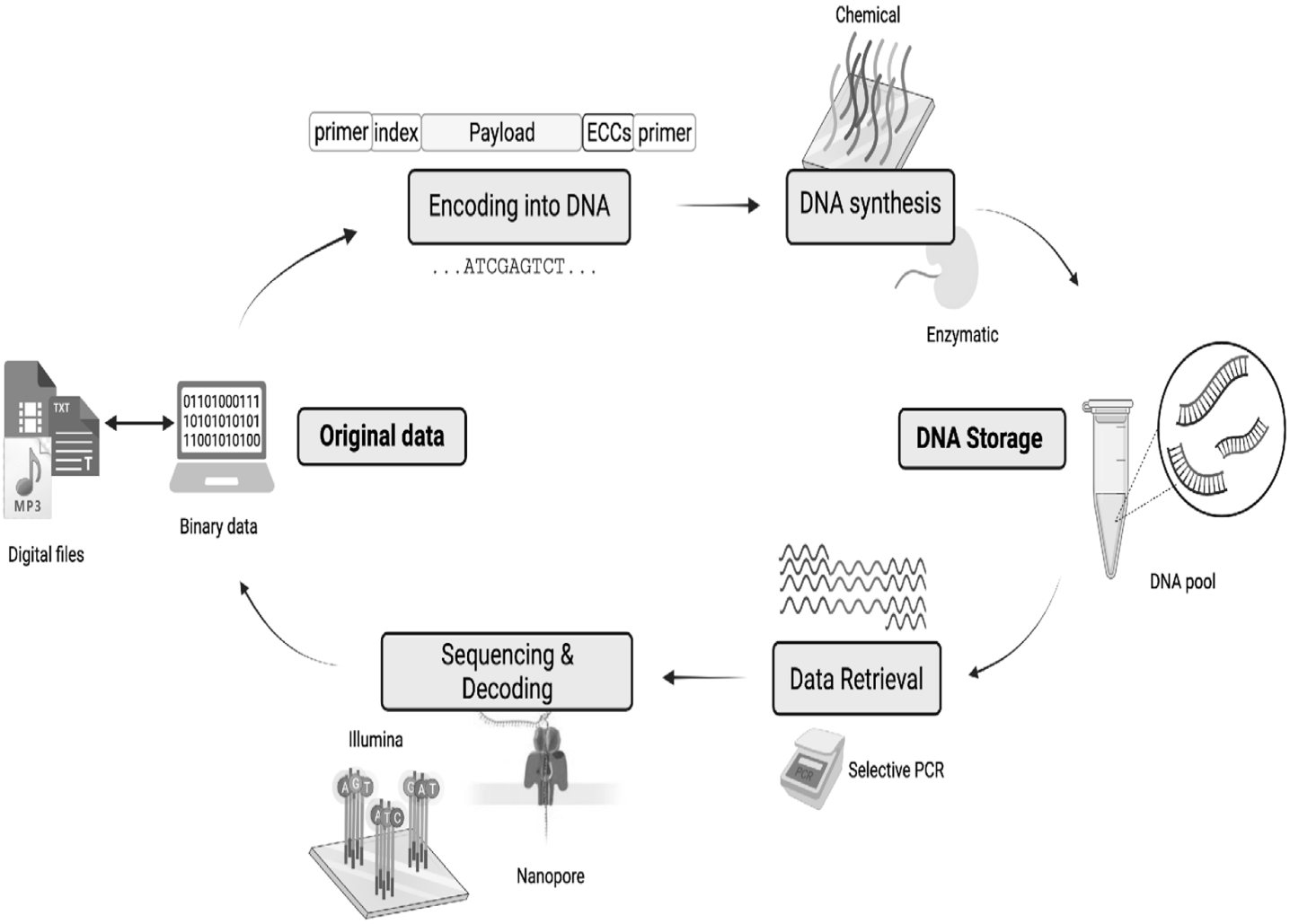
DNA is the solution

Yes, recently scientists have recovered, the oldest, two-million-year-old DNA sequences from frozen soil in the north-eastern tip of Greenland (Kjær *et al*., 2022). Which suggests the possibility of storing the information for more than 2 million-years provided with proper storage condition. When comes to the storage capacity, DNA possesses a high storage density. It is possible to store 215 petabytes (215 million gigabytes) of information in a single gram of DNA (Service, 2017). Any data can be stored within the eco-friendly DNA sequence and can be used in steganography. Data retrieval from the stored in DNA is relatively easy and rapid through polymerase chain reaction (PCR) and sequencing techniques. The concept was developed in 1960s. DNA digital system contains four letter discrete codes (A, T, G and C), which has the theoretical coding potential to store at least twice as much information in quaternary scaffolds as binary codes. Different binary transcoding methods like, Simple coding scheme, Huffman coding scheme, Improved Huffman coding scheme, Galois field and Reed-Solomon code based coding scheme, Forward error correction coding scheme and Fountain code - based coding schemes (Ezekannagha *et al*. 2022) can be used to code the nucleotides with the binary codes, zeros and ones .

Steps involved in the process

Mainly there are 6 steps in the entire process of DNA data storage. (1) Encoding of the digital information into DNA sequences, (2) data writing (synthesis of new oligonucleotides), (3) storing the synthesized DNA sequence in physical or biological conditions, (4) random access, (5) data readout via DNA sequencing, and (6) decoding the DNA sequences back into the original digital code, (Doricchi *et al*., 2022). The information to be stored is encoded in the DNA segment, which is called as information DNAs (iDNAs). Initially the original digital files will be converted into binary data (0s and 1s), further encoded into DNA sequences using nucleotide bases. Currently the technology is limited to short DNA molecules or oligos of about 300 nucleotides to maintain high degree of precision. The iDNAs are flanked by unique set of polymerase chain reaction (PCR) primer sequences (both forward, F and reverse, R) which acts as a key to retrieve the sequence information when it is in need. These synthetic oligos can be synthesized either chemically or enzymatically. Chemical synthesis is usually performed using either traditional column-based synthesis or array-based synthesis which is based on phosphoramidite chemistry methods. Enzymatic synthesis has relied on the Terminal deoxynucleotidyl Transferase activity of DNA polymerase. Once after the synthesis the DNA sequences will be stored by chemical (by encapsulation with silica nanobeads in microplates) or physical (in stainless steel storage capsules with glass coating inside) methods. The stored information can be retrieved using PCR and sequencing technologies. The encrypted message in the stored

oligo sequences will be decoded to obtain the original binary data and the original data files are recovered at the end.



Overview Process of a synthetic DNA-based storage system. (Source: Ezekannagha *et al.* 2022)

CONCLUSION

In conclusion, the process of using synthetic DNA as a storage medium offers great opportunity to meet the task of long-term archiving. Due to it's ultra-compactness, hyper density, ultra-persistence and massive redundancy the method has certainly ticks all the solution boxes for today's storage problems. This can also be considered as green data storage medium as it is eco-friendly in comparison to conventional storage technologies.

REFERENCES

Doricchi, A., Platnich, C. M., Gimpel, A., Horn, F., Earle, M., Lanzavecchia, G., Cortajarena, A. L., Liz-Marzán, L. M., Liu, N., Heckel, R., and Grass, R. N. (2022). Emerging Approaches to DNA Data Storage: Challenges and Prospects. *ACS nano.* 16(11): 17552-17571.

Ezekannagha, C., Becker, A., Heider, D., and Hattab, G. (2022). Design considerations for advancing data storage with synthetic DNA for long-term archiving. *Materials Today Bio*.100306.

Kjær, K. H., Winther Pedersen, M., De-Sanctis, B., De-Cahsan, B., Korneliussen, T. S., Michelsen, C. S., Sand, K. K., Jelavic, S., Ruter, A. H., Schmidt, A. M., and Kjeldsen, K.K. (2022). A 2-million-year-old ecosystem in Greenland uncovered by environmental DNA. *Nature*, 612(7939): 283-291.

Limbachiya, D., Gupta, M. K., and Aggarwal, V. (2022). 10 Years of Natural Data Storage. IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 8(4): 263-275.

Service, R. F., (2017). DNA could store all of the world's data in one room. *Science.* 15