ARTICLE

Digital information storage on DNA in living organisms

Authors
Jian Sun, Qiming Wang[1] (First authors)
Wenyi Diao, Chi Zhou, Bingbing Wang, Liqun Rao, Ping Yang

Affiliation
Synbio Technologies LLC, 1 Deer Park Drive, NJ, 08852, USA
[1]College of bioscience and biotechnology, Hunan Agricultural University, Changsha, 410128, Hunan, China

***Corresponding author:**
Jian Sun
jian.sun@synbio-tech.com

Abstract

The growing demand of digital information storage worldwide has led to the development of technology of using DNA as novel storage media. DNA is a suitable storage method due to its high data density, environment compatibility and long-term storage potential. Currently, most studies on DNA storage are based on short oligonucleotide pool synthesized on silica chip. However, despite the low cost and high-throughput advantages, this type of DNA storage also has shortcomings such as limited DNA quantity, difficulty replicating, etc. Thus, a new type of storing digital information within the DNA of living organism is attracting more attentions. In this research we conducted pilot studies of DNA storage in representative living organisms such as E. coli, yeast and Arabidopsis. This study aimed to address fundamental questions of DNA storage in living organisms, such as feasibility, stability and so on. From this study, we found that digital information can be stored and stably transmitted on DNA within these living organisms.

Introduction:

In the past decade, the global demand of digital information has grown rapidly. The total information storage quantity has exceeded 2.7ZB while consistently increasing by 50% each year[1], and expected to reach $3x10^{24}$ bits by 2040[2]. The data storage media basing on semi-conductive silica material currently being used can hardly keep up with this growing speed and demand, making a more dense data storage media with prolonged persistence time highly desired[3]. As an informatics molecule, the chemical properties of deoxyribonucleic acid (DNA) make it an excellent candidate material for storing digital information [4, 5]. When compared with current major data storage media, DNA has significant advantages of high data density, prolonged storage time along with carbon-based, environmental protection[6]. It is estimated that DNA storage media has the capability of storing all the data currently present on internet within a device smaller than just one cubic inch[4]. Moreover DNA can be packed in specific format [7] capable of resisting hostile conditions such as high temperatures, humidity and exposure to magnetic fields while preserving the data stored within. In addition to these advantages, DNA is also environment friendly and can be easily degraded and recycled in nature, causing almost no environment burden[8]. All of these merits endow DNA with the potential of becoming next generation of information storage media, especially for immutable, high-latency, sequential access application such as archival storage[9].

On the other hand, although storing digital information in DNA is very attractive to many researches, before it becomes a practicable and affordable method, many technique barriers and cost need to be well solved or overcome. These barriers include effective encoding strategy[10], high-throughput, low-cost synthesis[11], error correction mechanism, indexing, random access data retrieval method [12], efficient DNA sequencing techniques and data decoding procedures[13]. Many studies by different researchers using different strategies have been conducted on information storage in DNA and all have their collective advantages and disadvantages[14].

Before digital data can be stored within DNA, the information needs to be converted into the corresponding sequence made of the four DNA nucleotides A, T, G, and C. The major digital data we are currently using are encoded and stored as binary digits 0 and 1. So, ideally one base nucleotide represents 2 bits of binary information, thus are twice compressed over binary data. Considering indexing, error-correction and avoiding non-synthesizable sequences, an encoding efficiency of more than 1.6 would be a good encrypting algorithm [15, 16]. In this study we used our own developed and patented encoding algorithm and software (DNA Studio®) to encode a famous Chinese poem, 153 characters in length, into a nucleotide sequence with an overall coding efficiency of 1.98 (153x16bits/1232bp).

Following the encoding of digital data into a DNA sequence, the chemical molecule of DNA needs to be synthesized for physical storage. Current studies and practices of these synthesized DNA are mainly in format of short, single-stranded oligonucleotide pools synthesized by high-throughput, automated DNA synthesis microarray [11, 17-19]. These strategies have advantages of being fast, highly-automated and generating data that is easily retrieved. But they also have challenging disadvantages. Firstly, there are errors more

frequently occurring due to nucleotide mis-incorporation during chemical synthesis. Due to the limited quantity of each synthesized oligonucleotide presenting in a pool, a sophisticated and effective error correction mechanisms need to be employed [19, 20]. Secondly, due to short length of each synthesized oligonucleotide a large number of overlapping regions and complex indexing criterial to are necessary to encode long information [21, 22]. This leads to an increase in the overall cost and time to synthesize the accurate sequence. Thirdly, the synthesized DNA material is limited, a typical yield of synthesized DNA on an inch by inch size chip ranges from nanograms to micrograms, and the data reading (sequencing) process is DNA material consuming. Thus the number of data reading is limited and costly. Additionally it is difficult to replicate the original DNA sequence leading to further downstream issues. Fourthly, these types of DNA information storage are not essentially energy free and environment green, as long-term preservation of DNA harboring stored data requires low temperate, thus consuming additional electricity.

Compared with data storage in form of synthesized oligonucleotide pools, there were considerations and attempts to store the data DNA in living organisms such as bacterium [23, 24]. Storing data within DNA in living organisms has several advantages and overcomes many disadvantages associated with synthesized oligonucleotide pool. The major advantages are that living organisms can harbor very long DNA pieces, over a million base pairs in yeast for example, thus increasing effective coding DNA ratio. The living organisms can propagate very easily, giving rapid, unlimited and almost costless replicating ability. Also, in living organisms the informatics DNA is double-stranded which increases its stability. If further packed into some kind of environment resistance organisms such as spore, the DNA is expected to keep intact over one hundred years or even longer without special care[7].

In this study, we transformed the data DNA into E. coli, yeast and Arabidopsis, in order to see if the data DNA could be stored in these living organisms and how the data behaves in living organisms. In yeast and arabidopsis, the data DNA sequence were inserted into the host genome and the transgenic lines were generated. The data DNA sequences were then amplified out again from their progenies and were sequenced to check if any mutations occurred. By this study, we are hoping to examine the potential and feasibility of storing data within the DNA of living organisms, thus providing pilot information for this new type of potential long-term DNA storage technology.

Besides DNA storage media, a rapid, quick and random data retrieval mechanism must be developed for an applicable and effective data storage system[25]. Some research has been done on mechanisms for indexing, barcoding or sorting the synthesized DNA fragments[12]. Depending on the storage strategy, this data indexing or retrieval process could be simple or complicated. It can be complicated for DNA storage based on short oligo pools [22, 26] but much simpler for DNA storage by large DNA pieces. In this study, we are hoping to explore the possibilities of storing large DNA pieces rather than short sequences. This approach allows us to try to simplify the data reading process, because living organisms such as yeast it have the potential to store large pieces of DNA more than one million base pairs length. There were also studies to develop rapid and

random access to living DNA storage archive system by naturally-occurring biological procedures[27], paving the way to high efficient data retrieval of DNA storage.

Finally, DNA sequencing is still essential for reading the DNA sequence to retrieve the encoded information. This step developed much faster than other stages within the DNA storage process. Currently high-throughput, cost effective large-scale parallel DNA sequencing platforms are commercially available. There is also a DNA sequencing platform with extreme long reading capability under development within the synthetic biology industry. If this sequencing platform is available in the future, it will make DNA storage in living organisms more applicable and advantageous. Finally, DNA fragment overlapping assembly and DNA decoding programs need to be developed to account for the specific overlapping design and encoding scheme[26].

In this study, we tested the possibility of storing DNA in living organisms along with the stability of stored DNA. Although it is only a small part of developing a complete DNA storage system, it sheds light on a new mechanism of achieving DNA storage through an alternative method other than using synthesized oligonucleotide pools.

Materials and Methods:

1. Generate DNA sequence to be stored

A Chinese poem, “Snow”, was inputted into text file on a computer and was then saved as utf-8 code. The file was uploaded onto Synbio Technology’ internal server for converting to a DNA sequence. Two sequences were generated by DNA Studio version 1 and version 2 which were then downloaded. The two sequences are 2448 bp and 1232 bp long respectively (sequence 1 and 2 in appendix).

DNA Studio is a computer program that converts binary digital data of “0’ and “1” to letters of “A”,“T”,“G” or “C”, which represent the four different bases of DNA. The sequence of the four bases can be synthesized by a chemical and biological procedure of DNA synthesis and thus the original digital information is stored into DNA as a format of linear sequence of four different bases.

This is the how DNA studio encodes and decodes the digital data. A standard library between the ASCII chart and DNA sequence was built and one base corresponds to one binary bit. Thus the 153 character poem of “Snow” was converted to a 2448 base sequence (one Chinese character occupies 16 bits of binary data). In an improved version (version 2) of DNA studio, an optimized algorithm is used and the output DNA sequence is much shorter than version 1. Additional information about this procedure may be found in one of our previous publications[28].

Many text data have been tested on DNA Studio. They were encrypted in DNA sequence. The resulted DNA sequences were then decoded using decoding function of DNA Studio. All of them can be correctly decoded to give the exact same texts as input, indicating the well function and validity of the program. An online web interface of DNA studio is also available at https://www.synbio-tech.com/dna-studio/.

2. Synthesize DNA encoding “Snow”

DNA of “Snow” was synthesized with the overlapping PCR method from chemically

synthesized short oligo nucleotides. Fifty-eight oligonucleotides, with an average length of 65nt, were synthesized using (machine model, Synbio Technologies). The oligos were then assembled and ligated using the overlapping PCR method to produce the full length designed DNA[29].

3. Store the DNA of "Snow" in E.coli and Yeast

The above synthesized "Snow" DNA was then sub-cloned into the pUC57 vector and transformed into E. coli strain T1. The transformed E. coli bacteria were then selected by ampicillin resistance and grew over ten generations. One generation means overnight culture from bacterium seed to 4ml culture in LB medium. Plasmid was extracted out of the bacteria and the "Snow" sequence was sequenced.

The "Snow" DNA was sub-cloned into yeast strain *Pichia pastoris* vector pPIC9K by EcoRI and NotI. The Amp resistant clones were selected and the insert "Snow" DNA was sequence verified. Then the plasmid was linearized by AscI and was used to transform *Pichia pastoris* competent cell by electroporation (Bio-rad, GenePulser), using Bio-rad recommended parameters for yeast transformation. After pulse, the cell was incubated with 1M sorbitol for 1h and then plated onto MD plate (1.34g/L yeast nitrogen base, 2g/L Glucose, $4X10^{-4}$g/L biotin, 20g/L agar) for growth in 30°C for 2-3 days. Five positive clones were selected and seeded into 4ml YPD medium. After growing in 30°C shaker for 36-48h, 1 ml yeast cell was centrifuged, and the genomic DNA was extracted using yeast a genomic DNA extraction kit. Then the "Snow" DNA was amplified out of the genomic DNA and sequenced to verify for accuracy.

To test the DNA stability in yeast cell, we cultured the yeast cell harboring "Snow" DNA for ten cycles, from seed to 4ml liquid culture. The genomic DNA was extracted again, and the stored DNA was then amplified and fully sequenced. This obtained sequence was then compared to the original sequence to detect any changes.

4. Generation transgenic arabidopsis plants

The "Snow" DNA was sub-cloned into binary vector pDT1 and transformed into agrobacteria. The agrobacteria was used to transform Arabidopsis line Columbia by the flower dipping method[30]. The seeds from the transformed flowers were germinated and when the seedlings grew 3-5 leaves they were sprayed with herbicide. The transformed seedlings were then selected (Figure 4). The integration of "Snow" DNA into Arabidopsis genome was confirmed by PCR amplification and subsequent sequencing of target bands of expected size (Figure 5).

5. Decoding the stored "Snow" DNA

The DNA sequences retrieved from E. coli, yeast and Arabidopsis were decoded to its original digital information using DNA Studio. First the DNA sequences were saved as text files and uploaded onto Synbio Technologies' server. Then the decoded files were downloaded. Finally this digital information was compared to its original input.

Results:

1. Conversion of digital data into DNA

Several encoding algorithm to convert digital data to DNA sequence were previously developed such as the Huffman code, the

comma code and the alternating code [1, 15]. Each algorithm can efficiently encode digital data to DNA sequences but also have some limitations. To circumvent these limitations, we designed our own encoding algorithm and developed a new DNA encoding and decoding platform "DNA Studio®". This algorithm is capable of converting binary data into DNA sequences and also able to retrieve binary data back from DNA sequence. In this study, we chose a famous poem, "Snow", of Chinese Chair Mao as the information to be stored in DNA. This poem contains 153 Chinese characters including punctuations. Since one Chinese character occupies two bytes, or 16 bits of binary code, 153 Chinese characters occupy 2448 bits collectively. Previously our DNA Studio version 1 was recently upgraded to version 2, but we used both versions of DNA Studio to encode this text into DNA sequences. The version 1 outputted a DNA sequence of 2448 base pairs and the version 2 outputted a much shorter sequence of 1232 base pairs with few repeated sequences embedded within. Thus, the coding density or efficiency are 1 (2448 bits/2448 bps) and 1.98 (2448 bits/1232 bps). The original text and both DNA sequences are listed in appendix as figure 1, sequence 1 and 2. In this study, only sequence 1 was synthesized for further study.

2. DNA synthesis of data DNA

DNA sequence 1 was synthesized using a PCR based oligo overlapping assembly approach by Synbio Technologies. Fifty-eight oligo nucleotides were synthesized and assembled into the full length 2448 bp DNA sequence. The sequence was then cloned into the pUC57 vector and transformed into E. coli cell. Single-colony plasmid was extracted and sequenced to verify the accuracy of synthesized DNA.

3. DNA storage in E.coli

To test the stability of DNA stored in E.coli, the bacteria harboring the synthesized DNA was grown more than ten times. One time means that the bacteria was used as seed and 0.1% was inoculated into new LB medium and cultivated overnight. The plasmid DNA was then extracted, and the synthesized gene was sequenced for verification. We sequenced ten single-colony clones and found no mutation compared to the original sequence, indicating the stored DNA was stable within these ten generations.

4. DNA storage in Yeast

To test the stability of the DNA in yeast, we sub-cloned the synthesized DNA into the pPIC9K vector. Then the pPIC9K vector containing the synthesized DNA was linearized and transformed into a yeast strain Pichia pastoris. The yeast clone that successfully integrated the synthesized DNA into its genome was selected by selective medium and stored (Figure 2). The integrated DNA was then PCR amplified and sequence verified (Figure 3).

Figure 2: Transgenic yeast clones were grown and selected on MD plate.

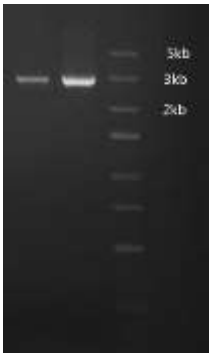


Figure 3: Integration of "Snow" DNA in yeast genome was confirmed by genomic PCR and sequencing

The yeast cell that contains "Snow" DNA was cultured for ten cycles. One cycle is defined as 36-48h cultivation from seed to liquid culture. Then the DNA was amplified again by genomic PCR and sequenced. Sequence alignment shows the exact same DNA sequence with no sequence change was found in ten individual clones.

5. DNA storage in Arabidopsis

DNA sequence encoding "Snow" was transformed into Arabidopsis with the flower dipping method and the transgenic seedlings were successfully selected by herbicide (Figure 4). Four primer pairs targeting different regions of the DNA sequence were used to confirm the integration of "Snow" DNA into the arabidopsis genome. At least three of five tested lines successfully amplified all of the four expected size bands from leaf the genome DNA sample, indicating the successful integration of target DNA sequence (Figure 5).

Figure 4: F1 generation transgenic seedlings were sprayed with herbicide. Only seedlings that were inserted with T-DNA containing "Snow sequence and *bar* gene could grow.

Figure 5: Five herbicide resistant seedlings were PCR tested using leaf genomic DNA. At least three produced all four bands of expected size.

To test if the insert DNA can keep intact during plant growth, we let the plant grow for an additional four weeks to produce more leaves. Then genomic DNA was then extracted from three individual lines and primers were used to amplify the whole gene. All of the three samples produced expected size bands. Finally the PCR products were sequenced and aligned with its original sequence. A 100% match was obtained in all of the three lines.

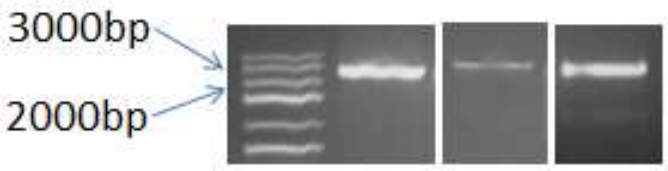


Figure 6: PCR amplification of "Snow" DNA from leaf genomic DNA, producing the expected 2.8kb band.

Discussion:

In this study, we encoded a small piece of digital information into a DNA sequence which was then synthesized *in vitro*. In order to investigate the possibility of storing DNA in a living organism we introduced the DNA into E. coli, yeast and Arabidopsis respectively. The feasibility and stability of DNA storage in these living organisms were then examined. Our preliminary results suggested that storing digital information within the DNA of a living organisms is both feasible and achievable.

The DNA sequence we inserted into living organisms was encoded by version 1 of DNA Studio. This version has more short repetitive regions and more unbalanced GC content (32.9%) than that of version 2 of Studio®. We chose this sequence because this complexity could be more challengeable to living organisms. If this sequence can be successfully stored, the other less complex sequences could be easily stored.

To our knowledge, successfully storing "Snow" DNA in E. coli may be not surprising. This is because E. coli is a simple, well-studied organism with many examples of successful cloning or integrating different kinds of DNA sequences into its genome. Plus that E. coli cell is not recombination efficient nor does it have a highly dynamic genome, making the successful storing of this "Snow" DNA rational and obtainable.

However, this situation is different when working with the yeast genome. Compared to E. coli cell, yeast generally has a much more dynamic genome. It has efficient a recombination mechanism that could exchange DNA pieces within its genome as well as between the genome and incoming exogenous DNA. This property of yeast cell is a potential concern for DNA storage, as stored information DNA may be unexpectedly altered. On the other hand, this property could be interesting to us because it may imply a mechanism or pattern that yeast cell deals with this kind of exogenous DNA. This can give us insight on how digital information DNA is incorporated into the yeast genome. Since we only looked at limited generations and limited progenies the result showed no variation but that doesn't necessarily mean that no DNA sequence change occurred. For DNA storage purpose, as long as the frequency of this change is kept under

certain level, it would not pose a big issue to DNA storage application.

Storing digital information DNA in higher organisms, such as plants, can be more challenging because their cells are more complicated than single cell organisms. For example, large scale transformation and transgenic line generation is much difficult than single cell organisms. In this study, we did this pilot experiment of storing information DNA in Arabidopsis to explore the feasibility and stability of this approach. Based on our results, our digital information DNA was successfully stored in Arabidopsis and passed down to its F1 progeny without any alternations. Since the Arabidopsis genome DNA was extracted from a randomly selected leaf we assume that all of the cells contain the exact same copy of the DNA. During the growth of the Arabidopsis seedlings, this DNA was replicated to generated millions or even billions of copies, overriding the copy number of any printed artworks to date.

Another important aspect of storing DNA in higher organisms is that some higher organisms can form special kind of tissue that is much more environmentally tolerant than normal organisms, such as a spore. This tissue has the capabilities of facilitating long-term DNA preservation.

In this study, we only tested a limited number of clones and progeny seedlings. The fact that no variation was found doesn't necessarily exclude the possibility of a rare frequency sequence change events. For further studies, large-scale sequencing may be employed to investigate the potential sequence alternation and the pattern the host cell deal with this stored DNA.

References:

[1] P.Y. De Silva, G.U. Ganegoda, New Trends of Digital Data Storage in DNA, Biomed Res Int 2016 (2016) 8072463.

[2] D. Panda, K.A. Molla, M.J. Baig, A. Swain, D. Behera, M. Dash, DNA as a digital information storage device: hope or hype?, 3 Biotech 8(5) (2018) 239.

[3] C. Bancroft, T. Bowler, B. Bloom, C.T. Clelland, Long-term storage of information in DNA, Science 293(5536) (2001) 1763-5.

[4] M. Castillo, From hard drives to flash drives to DNA drives, AJNR Am J Neuroradiol 35(1) (2014) 1-2.

[5] B. Zakeri, T.K. Lu, DNA nanotechnology: new adventures for an old warhorse, Curr Opin Chem Biol 28 (2015) 9-14.

[6] F. Akram, I.U. Haq, H. Ali, A.T. Laghari, Trends to store digital data in DNA: an overview, Mol Biol Rep 45(5) (2018) 1479-1490.

[7] J.P. Cox, Long-term data storage in DNA, Trends Biotechnol 19(7) (2001) 247-50.

[8] M.R. Gillings, M. Hilbert, D.J. Kemp, Information in the Biosphere: Biological and Digital Worlds, Trends Ecol Evol 31(3) (2016) 180-189.

[9] G.M. Church, Y. Gao, S. Kosuri, Next-generation digital information storage in DNA, Science 337(6102) (2012) 1628.

[10] Y. Erlich, D. Zielinski, DNA Fountain enables a robust and efficient storage architecture, Science 355(6328) (2017) 950-954.

[11] S. Kosuri, G.M. Church, Large-scale de novo DNA synthesis: technologies and applications, Nat Methods 11(5) (2014) 499-507.

[12] S.M. Yazdi, Y. Yuan, J. Ma, H. Zhao, O. Milenkovic, A Rewritable, Random-Access DNA-Based Storage System, Sci Rep 5 (2015) 14138.

[13] R.D. Andersen, G. Bristol, Simplified reading and storage of DNA sequence data, Am Biotechnol Lab 9(4) (1991) 26.

[14] C. Mayer, G.R. McInroy, P. Murat, P. Van Delft, S. Balasubramanian, An Epigenetics-Inspired DNA-Based Data Storage System, Angew Chem Int Ed Engl 55(37) (2016) 11144-8.

[15] G.C. Smith, C.C. Fiddes, J.P. Hawkins, J.P. Cox, Some possible codes for encrypting data in DNA, Biotechnol Lett 25(14) (2003) 1125-30.

[16] C.M. Gearheart, B. Arazi, E.C. Rouchka, DNA-based random number generation in security circuitry, Biosystems 100(3) (2010) 208-14.

[17] B. Hwang, D. Bang, Toward a new paradigm of DNA writing using a massively parallel sequencing platform and degenerate oligonucleotide, Sci Rep 6 (2016) 37176.

[18] N. Goldman, P. Bertone, S. Chen, C. Dessimoz, E.M. LeProust, B. Sipos, E. Birney, Towards practical, high-capacity, low-maintenance information storage in synthesized DNA, Nature 494(7435) (2013) 77-80.

[19] A.K. Yim, A.C. Yu, J.W. Li, A.I. Wong, J.F. Loo, K.M. Chan, S.K. Kong, K.Y. Yip, T.F. Chan, The Essential Component in DNA-Based Information Storage System: Robust Error-Tolerating Module, Front Bioeng Biotechnol 2 (2014) 49.

[20] R.N. Grass, R. Heckel, M. Puddu, D. Paunescu, W.J. Stark, Robust chemical preservation of digital information on DNA in

silica with error-correcting codes, Angew Chem Int Ed Engl 54(8) (2015) 2552-5.

[21] Q. Xu, M.R. Schlabach, G.J. Hannon, S.J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes, Proc Natl Acad Sci U S A 106(7) (2009) 2289-94.

[22] L. Organick, S.D. Ang, Y.J. Chen, R. Lopez, S. Yekhanin, K. Makarychev, M.Z. Racz, G. Kamath, P. Gopalan, B. Nguyen, C.N. Takahashi, S. Newman, H.Y. Parker, C. Rashtchian, K. Stewart, G. Gupta, R. Carlson, J. Mulligan, D. Carmean, G. Seelig, L. Ceze, K. Strauss, Random access in large-scale DNA data storage, Nat Biotechnol 36(3) (2018) 242-248.

[23] N. Yachie, Y. Ohashi, M. Tomita, Stabilizing synthetic data in the DNA of living organisms, Syst Synth Biol 2(1-2) (2008) 19-25.

[24] J. Bonnet, P. Subsoontorn, D. Endy, Rewritable digital data storage in live cells via engineered control of recombination directionality, Proc Natl Acad Sci U S A 109(23) (2012) 8884-9.

[25] J. Bornholt, R. Lopez, D.M. Carmean, L. Ceze, G. Seelig, K. Strauss, A DNA-Based Archival Storage System %J SIGPLAN Not, 51(4) (2016) 637-649.

[26] S.A. Tsaftaris, A.K. Katsaggelos, Retrieval efficiency of DNA-based databases of digital signals, IEEE Trans Nanobioscience 8(3) (2009) 259-70.

[27] F.G. TAVELLA, A. DOOLEY-CULLINANE, T. M. CONTI, M. COFFEY, L. BALASUBRAMANIAM, S., DNA Molecular Storage System: Transferring Digitally Encoded Information through Bacterial Nanonetworks, Cornell University Library Submitted (2018).

[28] Y. Zhong, S. Qi, F. Sheng, J. Tian, P. Zhu, P. Yang, X. Cai, A new digital information storing and reading system based on synthetic DNA, Sci China Life Sci 61(6) (2018) 733-735.

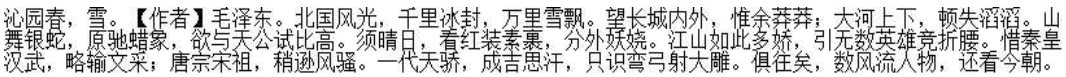
[29] G. Li, B.X. Dong, Y.H. Liu, C.J. Li, L.P. Zhang, Gene synthesis method based on overlap extension PCR and DNAWorks program, Methods Mol Biol 1073 (2013) 9-17.

[30] J. Schell, M. Van Montagu, The Ti-plasmid of Agrobacterium tumefaciens, a natural vector for the introduction of nif genes in plants?, Basic Life Sci 9 (1977) 159-79.

[31] F. Tavella, A. Marcinkevicius, F. Krausz, 90 mJ parametric chirped pulse amplification of 10 fs pulses, Opt Express 14(26) (2006) 12822-7.

Appendix:

1. Figure 1: Text information of “Snow” encoded into DNA



2. Sequence 1: DNA sequence encoding “Snow” created by “DNA Studio®” version 1

ATCACTGAACTAAATGAGATATCACGTACGATATCAATCAATACAGATCCATCCATAAG
TCTGAAGTAATCACGTACAGTATAGAAGTAAGTATACATAGAAGTAATGAAGTACATC
CATAGATCTGAACTAAAGTAAGTAGATATAGAAGTAATGAATGATCACTAGCGATCTA
GATCACTGACCATCGATACATCGTAAATGCTGAATAGAAGTAAGTATACAGATATAGA
ATGATGAAGATATCACCATCGATAGTAACTACTGACGTAAGATAATGACATAGTACCAT
CCATAAGTCTGAAGATATAGACATATAGAGTAAATGCTGACTGAAGATAATGCTAGAA
GTAGATCTGAAAGTAATGCCATCCATAAGTCTGAACATCGTAAAGTATGAAGTAAATG
CTGACTGAAGTAATCACGTACAGTAGTAACTACGATACTAATAGAAGTAAGTATACATC
AATGAAATGCTAGAGTAAGATATGACCATAGATATGACTGACGTAAGATAATGACTAA
GATAGATAGTAAATGATCACCATCCATAAGTCTGAATCAAAGTCGATCCATACATCCAT
AGATAGTAACTAATAGCTAGCGATACTAATAGCTAGCGATCCATCCATAATGCTAGAG
ATAGTAATACATGAATCACTGACTAGATAGACATCGTAAAGTCAGTACATCGTAAAGTC
TAGCCATCCATAAGTCTGAAGTAACTAATGACCATAGATAGTAATAGAATGATCACGTA
CGATACATATCACGTACGATACATATAGAAGTAAGTATACAGATCTGAATGAAATGAC
TAATACAATGCGTAAGTAACATCCATATCAACTAATCACTGAATGACCATCCATAAGTC
TGAAGATATAGAGTACCATAGTACAGTATGAAAGTACTAATGAATACAATGACTACTG
AATCAAATGCCATCCATAAGTCTGAATCACTAGATAGATACACATCGTAAAGTCGTAAG
ATAGTAATACAGTAAGATAATGATCACTGAACTACTAGCGATAGATATCACTAGCGAT
ACATAGTACAGTCGATACTAATAGAAGTAAGTATACAGTAACTAATGACTAGATCAAT
CAATGAACATATCAAGATCGTAAGATCCATCCATAAGTCTGAATGAATGAAAGTCTAG
ATGACGTACAGTATACACTAACTACTGAAGATATGACGATATACAAGTACTAACTACCA
TAGTACCATCCATAAGTCTGAAGATATACAAGTATCAAGATAGTAAATGATCAAGATA
GTAAGTAATCAAGATCAGTAAGTATCAATAGAAGTAAGTATACATCACTGAAGATCCA
TAGATCTGAATGAAATGAGATAGTAACTAATACATCACTAGATCAACATAGATAGTAA
ATGCAGTAGATCAGTAAGTATGACCATCCATAAGTCTGAAGATCCATAATGAGATATCA
AGATCGTAAAGTATCAAGATATGAAAGTACTAATACCCATAATGAGTAATCACTGAAC
ATATGACAGTCGATCGTAATCAATACAGTAACTAACTAAATGATGAAAGTATAGAAGT
AAGTATACATCAAAGTCGATCTGAATGAAGTACGTAATCAATGAATCAACTAATGAAT
CACTGAACATAGTAATCACTAGATCAATCACCATCCATAAGTCTGAATGAAGATATCAA
GATACTACCATAGTAATAGATCAAGATACTAATGAAGTAAATGCTGAATGACCATCCAT
AATGCTAGAGATAGATAATGAAGTAGATCTAGAGTAATGAAGATCTAGACTACTAGAT
GAAGTAAGATATCACCATCCATAAGTCTGAATGACAGTAAGTCGATAGTAAAGTAAGT
CAGTAGTAACTACTGACGTAAGTACAGTAGTACAGTATAGAAGTAAGTATACACATCG
TAAAGTAAGTACATCGTACGTAATAGAGATAGTAATACAGTAAGTACAGTACTAACAT
CCATCCATAAGTCTGAATCAATACAATGAAGTAGATACATAAGTAGTAATCAAAGTAA
TGCGATATCACTGAAGATATGACCATCCATAAGTCTGAAGATATAGCGTACAGTACTAC
TAGCTGAATCAAGATCCATATACCCATAGATCCATAATGATAGAGATCTGAAAGTACAT
AGATAGTAATACATGAAGTAATCACGATAGATATAGAAGTAAGTATACACATCCATCC
ATAATGAGATCCATACTAAAGTATGAATGACGTAATAGCCATCCATAAGTCTGAATCAA
GATATGAAAGTAGTAACTACTGACGTAATCACGATACATAATGACATCGTACTAGCAGT

ATGAATACATCAAGTACCATCCATAAGTCTGAACTACCATCGATACTAATGAATGAAAG
TCTAGACATCGTACTGACAGTATCAATGAAATGCGATATAGAAGTAAGTATAC

3. Sequence 2: DNA sequence encoding "Snow" created by "DNA Studio®" version 2

GGGCGGGGTCGACAATTTTCGCGTTCTCACTTGGGGAAGACTTCGCCCAGAAAAACAG
AAATAATAGGTTGACAAAAATTAGAAATATTCCGGTCGTCGAGGGTTAGCATGAAGAA
AAACTTAGATTGTTTCGGGTCTCAGAGCTTATTACTGGGGAAGATTAGTAAGCTATGAG
ATTATCGAATTGAAAATGGGGAAGATAGCAATGCTATGAGACTTCGCCCCTCAGTCAA
GAAAAACTCTGATCGCTTTTGGGTTTGGAGCTTATCATTTTCTATTCGGGGAAGATCAA
GTGGTAGGTTCTCAAGCGGTCAAGCGGTGGGGATCGTTCTACTGTCGACGAGTAGCAAC
CTAGCAACGGGGGAAGACTCATGGGTTCTAGATTCGCGTTATCGCGTTAAGAAAAACTT
GATGATCAACATGCCTTAGGTCCATCGATGGGGGAAGATTAGCTGGCTCCTGAACATGA
CATCAGATCATGGGGAAGATCCGAGACTAGCAAGCTTCTACCTTTATTCGACACGGTTT
TCCGGTTACTCCGTCAAGAAAAACCTCATGCGTCTCTGTATCTTGCTTGGGGAAGATGT
GAACGTGGCCCACCACAGATTTGGTACAACACAGGCTGGGGAAGATTACAATCTTCTAT
TCTTCTCTTCTTCCAATCAGAAAAACTCGATTGGTTGATGATTTCTCAACTCCGTCTATT
CTATCCTTCCAATGGGGGAAGATTGGATTTTCTTGCAATCTTTGAACAACGGATCTTCG
ATATGCCGTGCTCACCTCACAATTGAAAGAAAAACTCAAGTGATGCTGCTCTGTCCATG
TCGATACTTCCGTCTCGGGGAAGATGTTTCTTCAGGCTAGTCTTCATGCTATGATGGGG
GATCGTTTTATAATTCGCTTGTTCGCACGTGCTTTTCGGGGAAGATGCCAAGTCTAAAA
CCCTCAGAGCCTCCCTCCAGAAAAACTAGCAAAATAGCGCAGTTCTACCTCTCCCATAG
GGGAAGATCACATAATTTAAACTTCAAATGTTCGATTTGGGGGAAGATTAGGCCCCACG
GATCTTGGACGGTTGGATAGTTGAAATATTCTACTGCTTCGTTTAGAAAAACTAGGGGA
TTTGGCAAATGTGGCAGGGGGAAGATCTTTGAACTCAGAGCTCGTTAATTAGCCGCCTG
ACTCCTGGGGAAGACAGGGTCATGTGAACGTAGCGACCTCTGATGTAGAAAAAC

