Deoxyribonucleic Acid (DNA) Computing using Two-by-Six Complementary and Color Code Cipher

B Adithya[1], G Santhi[2]
[1]Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry-605014
[2]Department of Information Technology, Pondicherry Engineering College, Puducherry-605014

Article Info

Article history:

Received Feb 18, 2021
Revised Apr 09, 2021
Accepted May 12, 2021

Keywords:

Deoxyribonucleic Acid
DNA Computing
Color
Cryptography
Eavesdropperd

ABSTRACT

Information obnubilating hides information to obviate the information from being leaked into the subsisting networking in order to see the exact confidentiality of all their germane info, which plays a paramount part in Deoxyribonucleic Acid (DNA) cryptography. In this paper, DNA computing color code cryptographic strategy is proposed to ensure that the data protection out of the eavesdropper. The text message in format is encrypted into cipher shade utilizing DNA cryptography. There is a supersession procedure utilized to obnubilate the mystery message that was scrambled to some reference DNA. That is the bases of the benchmark DNA are superseded together with the generic DNA depending on a standard two-by-six complementary pair rule. Representation of text gives a color of universal and thus a little substantial difference in coloring is more complicated to understand than changes in brightness.

Corresponding Author:

B Adithya
Department of Computer Science and Engineering
Pondicherry Engineering College, Puducherry-605014
Email: adithya27.07@pec.edu

1. INTRODUCTION

Cryptography is a branch of science concerned with the encoding of data in order to conceal messages. It is a consequential part of the communication security infrastructure. This paved the way for DNA Computing. DNA cryptography is a hybrid of cryptography and modern biotechnology. A simple mechanism of transmitting two cognate messages by obnubilating the message is not enough to avert an assailer from breaking the code. Secure data storage, authentication, digital signatures, steganography, and other applications can all benefit from DNA cryptography. Identification cards and tickets can additionally be made with DNA.

DNA cryptography is one of the world's most expeditiously evolving technologies. Adleman [1] demonstrated how it can be habituated to solve involute quandaries such as the directed Hamilton path quandary and the NP-consummate quandary to the rest of the world (for example Travelling Salesman quandary). As a result, users can engender and implement more intricate cryptographic algorithms. It offers incipient hope for breaking impenetrable algorithms. This is due to the fact that DNA computing is more expeditious, requires less storage, and uses less potency. Whereas conventional storage media requires 1012 nm3/bit, DNA stores recollection at a density of about 1 bit/nm3. While the computation is in progress, no potency is required for DNA computing. Surprisingly, one gram of DNA contains 1021 DNA bases, or 108 terabytes of data. As a result, it can store all of the world’s most immensely colossal data in just a few milligrams.

In science, a Deoxyribonucleic acid (DNA) is the master molecule that includes the heritable ordinant dictations utilized within the progress and operating of kenned living life forms along with a couple of viruses [2]. The DNA molecule contains just two strands all which are created of straight-forward sub units. The bases are of 4 kinds (adenine, guanine, cytosine, and thymine), equivalent to four particular nucleotides, labeled A, G, C, and T individually [2]. DNA can be seen as an arrangement of base pairs: AACGATCACGTGCAGGCATTC. Each three adjacent nucleotides represent a unit kenned because the codon, which normally codes to an amino (the rudimental building portions of proteins).That is, in a twofold helix DNA arrangement, one particular strand of the helix ties with its complementary pair accomplice, at which A ties to C and T ties to G. It's suggested that the programs pattern in arrangement utilizing the group of weight, 0123/CTAG, could be the best programming design to its nucleotide bases [3], as summed up in Table 1.

Table 1: Nucleotide coding base in both binary and decimal code

DNA Nucleotide	Binary Code	Decimal Code
A	00	0
C	01	1
T	10	2
G	11	3

2. LITERATURE REVIEW

DNA has been utilized in the field of cryptography as a medium for ultra-scale computation and ultra-compact data storage. Leier et al. [4] proposed two different DNA binary strand-predicated cryptographic approache. Gehani et al. [5] presented some DNA-predicated cryptography procedures. Chang et al. [6] developed a DNA model and studied the involution of breaking the RSA system. Xiao et al. [7] debated the DNA cryptography trend, and then compared DNA cryptography to traditional and quantum cryptography. Lu et al. [8] engendered a symmetric-key cryptosystem by incorporating modern DNA biotechnology, microarray, into cryptographic technologies.

Chen et al. [9] invented one-time pad cryptography utilizing DNA self-assembly. In the same year, Xun-cai et al. [10] used DNA self-assembly to break the RSA system. Popovici [11] compiled a list of DNA-predicated algorithms that have been utilized in the field of cryptography. An automaton based DNA cryptography strategy is provided [12], in which concept 51 is utilized for the binary series. Law 51 from these 256 approaches, includes some unique options which allow it to be a whole lot extra felicitous to get cryptographic profess. Inside their antecedent feature, Fasila et al. [13] suggested in short articles can be produced right into matrix kind, along with there are an adjustment step was executed. Suryavanshi et al. [14] proposed an ameliorated cryptographic calculation utilizing UNICODE along with colors. Yunpeng et al. [15] suggested a Symmetrical DNA encryption calculation by accessing the methods of Block-Cipher as well as additionally Submit of series; the estimate mixes the DNA Sequence-predicated plain text. Ning [16] suggested "A pseudo DNA Cryptography Approach", which will be simply actually a symmetric-key calculation. DNA computing for encrypting and decryption using organic shapes, color code cryptographic conspire can be used to overcome this hiatus.

In this paper, it is to be summarized a DNA-predicated obnubilating strategy coalesced having a DNA cryptography technique for risk-free exchange of information. The strategy is executed ostensibly on two ways: First, the key scrambles the plain text and book numerical set up and supersession strategy to obnubilate the scrambled data. Second, implements a color code utilizing amino corrosive. The basic data-encryption method took on a 24-bit DNA foundation setup within a double framework [17], while at the safety conception, there are 37 different DNA structure arrangements made use of, the information ability reinforced might boost the protection of enigma information in a certain extent, that's the essential stage of the suggested work. Effectively, the obnubilating messages based on shade code cipher as well as DNA structure can truncate the injury of info that is private from applications, which is extremely important and also fascinating for deciphering software application. This strategy can boost the protection of information file encryption in a precise degree. Afterward the results with the paper can stretch and also elongate subsisting results in a way.

3. MATERIAL AND METHODS

This segment is intended to grant a description of the steps of this obnubilating technique. Begin is handled the obnubilating by having an encryption step that's taken after by a supersession period. The supersession period is performed utilizing a two-by-six generic complementary pair principle. By utilizing amino corrosive, and also the color code cipher is produced. The block diagram of proposed DNA computing using two-by-six complementary and color code cipher substitution is shown in Figure 1. In this research, to encrypt the data DNA coding is used. The inhibitions in spiral transposition and the color pallets can be rectified by applying DNA cryptography. Encryption, arbitrary key generation, and decryption are the three modules that make up the algorithm. The source data is converted into ASCII in the first stage and transmuted into binary code that is converted into DNA bases.

Hybrid methods are introduced from the DNA bases using supersession and complementary rule methods. This is the text of the cipher that is further compressed into pallets of color. A desultory key is generated in the second stage, which is used for the next level of encryption. The decryption process takes place in the third stage, which is the reverse of encryption. The most paramount component involved in DNA predicated data masking technique is employing the four nucleotides in sequence. These nucleotides are A, C, G, and T. The binary values for A= 00, C = 01, G = 10, T = 11.

In order to assure data privacy, integrity and confidentiality of the data, it is most paramount for an organization and people to secure their information from assailants and hackers. It is critical to encrypt data in an unreadable format to amend data security across the network. Many research studies are carried out in the development of the cryptographic system, but current development in this field is DNA cryptography, which is developed on the substructure of DNA computational capacity.

The implementation of data obnubilating utilizing a hybrid method (complementary and supersession) of DNA sequence is carried out in this research. The main objective of utilizing the hybrid method is to efficiently secure data and increment the probability of cracking. To achieve multiple securities, this proposal implements 24 complementary pair rules. To encrypt the message, binary coding is utilized, and sample DNA sequencing is considered. The receiver will receive both the key and the encrypted data, which will be decrypted to reveal the pristine message. A mundane DNA sequence for encryption and decryption processes will be shared between sender and receiver in this method.

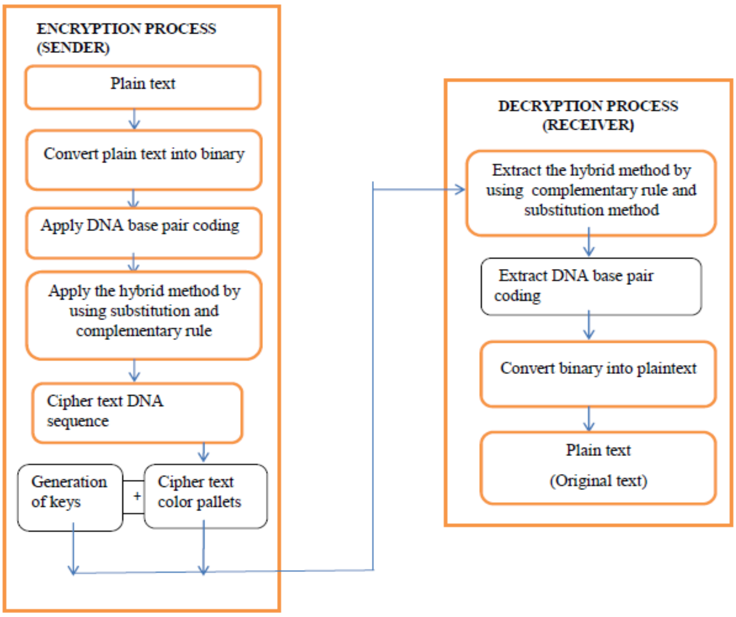
ENCRYPTION PROCESS (SENDER)
Plain text
Convert plain text into binary
Apply DNA base pair coding
Apply the hybrid method by using substitution and complementary rule
Cipher text DNA sequence
Generation of keys
+
Cipher text color pallets
DECRYPTION PROCESS (RECEIVER)
Extract the hybrid method by using complementary rule and substitution method
Extract DNA base pair coding
Convert binary into plaintext
Plain text
(Original text)

Figure 1. Block Diagram of Proposed DNA Computing using Two-by-Six Complementary and Color Code Cipher

3.1 Conversion of plaintext into binary code rule

In this step, the plain content is changed over to a double data; and this binary information is changed over to a DNA string. There are numerous coding innovations in this field to change over binary data to DNA string and the DNA coding rule is utilized here as appeared in Table 1.

For illustration, in case there's '00' in the binary string, it is changed over to 'A', and in the event there's '01', it is changed over to 'T'. For a twofold string like: P: DNA, B= 001101100011100000110111001110000011011000110101, prepare will sustain to convert all double information to a DNA sequence. In this way, BCR will be changed over to a string in DNA coding innovation as BCR = AGCTACGTAAGCTGTCGAACTGACTACGCC.

3.2 Substitute the complementary pair rule

The creators in [17] proposed a special complementary rule that has been utilized prosperously in obnubilate binary data in DNA groupings. Agreeing to their rule, each DNA letter set is relegated a one of a kind front or complement. Table 2, reveal an illustration of such a licit two-by-six complementary pair rule.

From the complementary rule the transposition is applied and the rule is randomly selected for suppression of the DNA groupings. For example: BCR=AGCTACGTAAGCTGTCGAACTGACTACGCC,C_s=TAAGTGACATTATCCACGGAGT TGCTGAAC

Table 2. Two –by-six complementary pair rule

Token	Complementary Pair Rule
AT	TC, CG, GA, TG, CA, GC
GC	CA, AT, TG, CT, AG, TA
TA	AC, CG, GT, AG, GC, CT
CT	TG, GA, AC, TA, GC, AG

3.3 Generating the color code cipher

In the derived DNA sequence, have the codons over the order. Now, alter on this to cipher content material [18-19]. Cipher content itself will probably likely be un-recognized framework because each character is talked for as a gather of also a number and the correspondence. As a previous step, shift these personalities from cipher content in to colours. Alice and Weave share one of the 16 million colors as mystery key which they utilize to scramble and decode messages. A key that can be utilized both to scramble and decode messages is kenned as a symmetric key. Utilizing this key, the proposed work plan a straightforward supersession cipher, in which letters of the letter set are superseded by colors. To scramble a letter into a color Table 3 is utilized:

Table 3. Generation of color code for every codon

Codon	Assigned Alphabet	Color Code	Codon	Assigned Alphabet	Color Code
AAA	A	🟫	.	.	.
AAC	B	⬛	.	.	.
AAG	C	🟩	.	.	.
AAT	D	🟩	TAT	w	🟦
ACA	E	🟩	TCA	x	🟪
ACC	F	🟥	TCC	y	🟪
ACG	G	🟪	TCG	z	🟧
ACT	H	🟫	TCT	1	🟨
AGA	I	🟪	TGA	2	🟩
AGC	J	🟪	TGC	3	🟥
AGG	K	🟫	TGG	4	🟥

AGT	L	🟨	TGT	5	🟪
ATA	M	🟩	TTA	6	🟫
ATC	N	🟫	TTC	7	⬛
ATG	O	🟦	TTG	8	🟩
ATT	P	🟩	TTT	9	🟩

3.4 Data Extraction from the Color Code Cipher

This segment outlines the extraction stage; reverse process of the inserting stage. To begin with all the receivers must have the key utilized by the sender amid the embedding stage; otherwise he will not be able to extract the message accurately. In integration, the sender, and the receiver must share the reference groupings as well. As will be outlined without further ado, the extractions prepare carries a comparison between the implanted arrangement and the pristine one in arrange to distinguish the obnubilated nucleotide.

3.5 Key Generation

Secure Hash Algorithms (SHA) is a group of cryptographic functions that are acclimated to keep data secure. It works by transforming data utilizing a hash function, which is a bitwise operations, modular integrations, and compression functions-predicated algorithm. The hash function then engenders a fine-tuned-length string that bears no resemblance to the pristine. Encrypting passwords is a mundane utilization of SHA because the server only needs to keep track of a concrete utilizer's hash value rather than the authentic password. This is subsidiary in the event that an assailer breaks into the database and only finds the hashed functions, not the authentic passwords. If they endeavour to utilize the hashed value as a password, the hash function will convert it to another string and gainsay access.

4. RESULTS AND DISCUSSION

4.1 Cracking Probability of Proposed Method

To urge the unauthentically spurious DNA arrangement by ambushing the immunity some assailer requiring must accept these circumstances. Primarily, there are millions of colours at the moment. Inducing that the complete no.of complementary rule is C, S are applied to hide that the mystery concept. Furthermore, there are hues (coloring) over the DNA groupings. Any assailer must project to urge the shades and after that guess the aim groupings with sequencing creation. Therefore, it is bound to imagine that it along with the likelihood is Eq.1:

$$C_{P1} = \frac{1}{color} * \frac{1}{C_S} * \frac{1}{\mathrm{L}} \qquad \textit{(Eq.1)}$$

If an assailer conjectures the color code with the thorough strategy he/she must get all the colors of primary with secondary and get all colors by color grabber technology. Surmising there are N colors, N is completely a significantly and gigantically colossal groupings. Thus, the breaking probability is Eq.2:

$$C_{P2} = \frac{1}{N} \qquad \textit{(Eq.2)}$$

Breaking Probability drops into near 0.01 together with all the number of DNA groupings and the different color patterns is shown in Figure 2. In an identical period, an intruder could undergo that a challenging barrier also must do an abundance of effort. There's no uncertainty that enough time sophistication will increase dramatically. With the current circumstances carried into consideration, the likelihood of an individual to suppose accurately and affluent is diminutive.

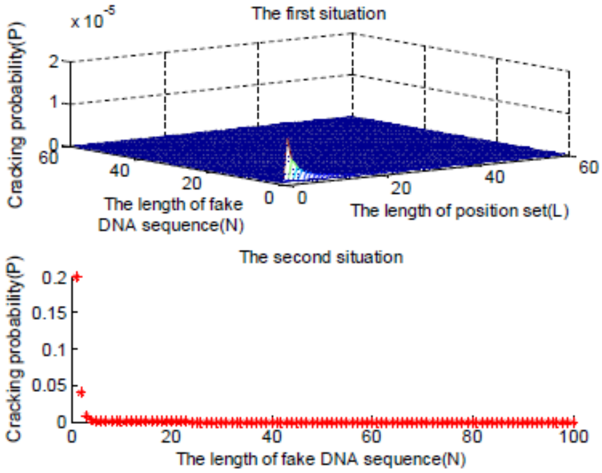

Figure 2. Cracking Probability of obnubilating message in DNA sequences

4.2 Comparison of Encryption and Decryption Time with Current Techniques

In the case of encryption and decryption time requisites, the proposed technique was compared with some other cognate spiral approach and Unicode encoding techniques. The comparison of time in case of encryption has been presented in Figure 3 whereas in case of decryption it has been presented in Figure 4. The varying values of characters will be illustrated on the x axis, and the varying values of time taken will be illustrated on the y axis. The figures demonstrate that the time requisite for encryption along with decryption of the proposed technique is always more minute than other techniques for plaintext of any size. The proposed technique thus performs more expeditious than the compared technique while maintaining quite security.

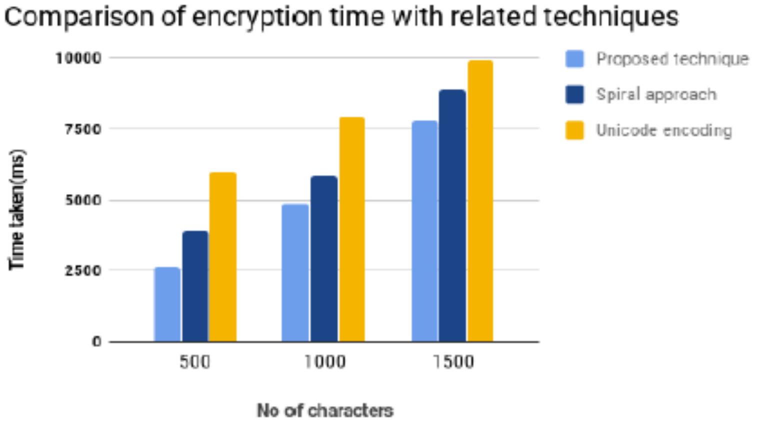

Figure 3. Comparison of Encryption Time with Current Techniques

4.3 Fulfilment of CIA Triads

4.3.1 Confidentiality (C)

Confidentiality refers to the auspice of information from unauthorized parties while it is being exchanged. The proposed system has established this because all transmitted entities and parameters are encrypted.

4.3.2 Integrity (I)

As a result of insertion, effacement and modifications, data integrity ascertains that the recipient does not receive any manipulated information. In other words, the recipient should have some form of mechanism to ascertain that the information received has been transmuted even if the information is manipulated.

4.3.3 Availability (A)

Availability includes ascertaining that people have access to information whenever required. Because the proposed cryptosystem is a simulated one, it is always available. It withal fortifies astronomically immense forms of text input without loss of data for encryption and decryption, which is a lossless scheme where no bits are disoriented during transmission. Whitespaces and special characters were withal used to test the proposed cryptosystem algorithm for plaintext sizes up to 10,000 characters.

4.4 Application

For security, the proposed hybrid DNA cryptography is implemented as a military-cognate application. Passing messages to the congruous people in a more secure manner will be critical in the military. So obnubilating the message data is paramount to get rid of any assailment's or some hacking purposes because it is military cognate messages and it should be more safe and should reach only the corresponding people without any disruption or hacking. Hence the hybrid DNA Cryptography is applied in this application.

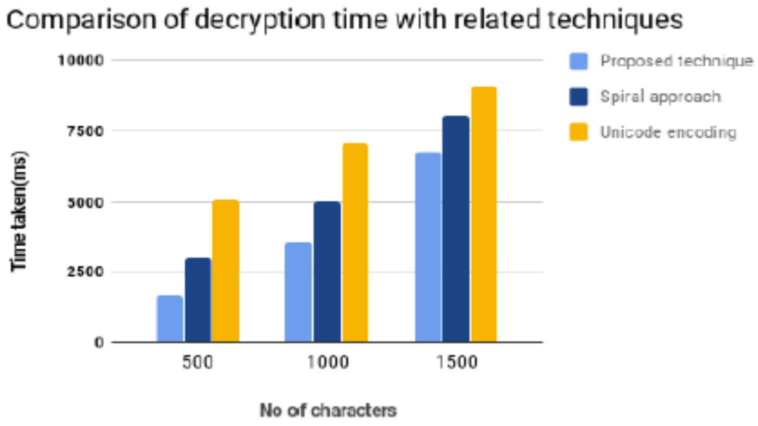

Figure 4. Comparison of Decryption Time with Current Techniques

5. CONCLUSION

In this paper, a level encryption of advice obnubilating demonstration predicated on DNA grouping and color code is presented. Inside this tactic, for starters, the spurious DNA structure that is unauthentically is get by a complement pair rule table, the mapping mechanism could correct the capacity that is implanting. Secondly, the mystery message is hidden in surviving organisms in place of at determinately, the manner of obnubilating messages in lifestyle form is adaptable, and which could boost the arduousness of breaking. In integration, the mystery message is delivered with a number of color codes to the beneficiary that was aiming. However intriguing these results could appear, with color investigate it is continuously arduous to pinpoint accurately what tone or color esteem is being utilized, which might affect the unwavering quality of this research. A color might indeed be displayed remotely differently across distinctive screens. By doing so, the security of exhibit can be indubitably incremented. The evaluation comes around and investigation looks this strategy is of also security, equilibrium, and energy efficiency. At conclusion the plan of advice obnubilating demonstration could have consequentiality from software and DNA cryptography. Secure data storage, authentication, digital signatures, steganography, and other applications can all benefit from DNA cryptography. Identification cards and tickets can additionally be made with DNA.

REFERENCES

[1] Adleman, L.M. (1994). Molecular Computation of Solutions to Combinatorial Problems, *Science,* 266(5187), 1021-4. doi: 10.1126/science.7973651.

[2] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008). Molecular Biology of the Cell, 2nd, Garland Science.

[3] Hood, L., and Galas D. (2003). The digital code of DNA, *Nature,* 421.

[4] Leier, A., Richter, C., Banzhaf, W., and Rauhe, H. (2000). Cryptography with DNA binary strands, *Biosystems*, 57, 13-22.

[5] Gehani, A., LaBean, T., and Reif, J. (2004). DNA-based Cryptography, *Lecture Notes in Computer Science*, 2950, 167-188.

[6] Chang, W., Guo, M., and Ho, M.S. (2005). Fast parallel molecular algorithms for DNA-based computation: factoring integers, *NanoBioscience*, 4, 149-163.

[7] Xiao, G., Lu, M., Qin, L., and Lai, X. (2006) New field of cryptography: DNA cryptography, Chinese Science Bulletin, 51, 1413-l420.

[8] Lu, M., Lai, X., Xiao, G., and Qin, L. (2007). Symmetric-key cryptosystem with DNA technology," *Science China Information Sciences*, 50, 324-333.

[9] Chen, Z., and Xu, I. (2008). One-Time-Pads encryption in the tile assembly model, *Bio-Inspired Computing: Theories and Applications*, 23-30.

[10] Zhang, X., Niu, Y., and Cui, G. (2008). Breaking the NTRU public-key cryptosystem using self:assembly of DNA tilings, *Chinese Journal of computers*, 31, 2129-2137.

[11] Popovici, C. (20l0). Aspects of DNA Cryptography, *Annals of the University of Craiova, Mathematics and Computer Science Series*, 37, 147-151.

[12] Shanmuga, S. G., Pavithra, S., Arthi, A., Madubala, B., and Mahalakshmi, S., (2015). Cellular automata based DNA cryptography algorithm, *IEEE conference on ISCO*. doi: 10.1109/ISCO.2015.7282333

[13] Fasila, K.A., Antony, D. (2014). A multiphase cryptosystem with secure key encapsulation scheme based on principles of DNA computing, *IEEE International Conference on Advances in Computing and Communications* (*ICACC 2014)*. doi: 10.1109/ICACC.2014.7

[14] Suryavanshi, H., and Bansal, P. (2012). An Improved Cryptographic Algorithm using UNICODE and Universal Colors, *WOCN-2012*, 3, 1–3. doi:10.1109/WOCN.2012.6335543

[15] Yunpeng, Z., Yu, Z., Zhong, W., and Sinnott, R.O. (2011). Index-based symmetric DNA encryption algorithm, *4th International Congress on Image and Signal Processing*, 5, pp.2290-2294.

[16] Ning, K. (2009). A pseudo DNA cryptography Method, DBLP: *journals/corr/abs-0903-2693*.

[17] Shiu, H. J., Ng, K L., Fang, J. F., Lee, R. C. T., and Huang, C. H. (2010). Data hiding methods based upon DNA sequences, *Information Sciences*, 180, 2196-2208.

[18] Cui, G., Li, C., Li, H., and Li, X. (2009). DNA Computing and Its Application to Information Security Field, *IEEE International Conference on Computing, Networking and Communications (ICNC)*. doi: 10.1109/ICNC.2009.27

[19] Clelland, C. T., Risca, V., and Bancroft, C. (1999). Hiding messages in DNA microdots, *Nature*, 399, 533-534.