



Information Hiding Based on DNA Sequences

Defaf Shiker Kadhum^{1*}, Sahar Adil Kadhum²

1 College of Sciences for Women, University of Babylon, defaf.kadim.gsci10@student.uobabylon.edu.iq, Hilla, Babel.

2 College of Sciences for Women, University of Babylon dr. sahar. Adill@gmail.com, Hilla, Babel, State.

*Corresponding author email defaf.kadim.gsci10@student.uobabylon.edu.iq; mobile: 07601014977

إخفاء المعلومات بالاعتماد على سلاسل الحمض النووي

ضفاف شاكرك كاظم^{1*}، سحر عادل كاظم²

1 كلية العلوم للبنات، جامعة بابل، defaf.kadim.gsci10@student.uobabylon.edu.iq، الحلة، بابل

2 كلية العلوم للبنات، جامعة بابل، dr.sahar.Adill@gmail.com، الحلة، بابل

Received: 4/8/2022

Accepted: 5/10/2022

Published:

31/12/2022

Abstract

Information security is a major source of worry, especially in light of the rapid expansion of internet use in recent years. As a result of this expansion, there have been incidences of illegal access, which have been mitigated by the adoption of a variety of secure communication protocols, including encryption and data concealment. DNA's bio-molecular properties have seen an uptick in popularity as a carrier for cryptography and data hiding in recent years. when information needs to be hidden. Therefore, DNA bases are utilized as information carriers in the data concealing strategy to increase safety. DNA-based steganography and cryptography combine a biological property with conventional methods to provide an algorithm with increased security. Because of their ability to maintain their chemical and biological characteristics, DNA sequences also have a high data capacity.

Keywords:

DNA, DNA Cryptography, DNA Steganography, SNP, Data Hiding.

الخلاصة

يعد أمن المعلومات مصدر قلق رئيسي، خاصة في ضوء التوسع السريع في استخدام الإنترنت في السنوات الأخيرة. نتيجة لهذا التوسع، كانت هناك حالات وصول غير قانوني، والتي تم تخفيفها من خلال اعتماد مجموعة متنوعة من بروتوكولات الاتصال الآمن، بما في ذلك التشفير وإخفاء البيانات. في السنوات الأخيرة، كانت هناك زيادة في استخدام الحمض النووي للتشفير وإخفاء البيانات كناقل، مع الاستفادة من قدراته الجزيئية الحيوية. في إخفاء البيانات. نتيجة لذلك، في نهج إخفاء البيانات، يتم استخدام قواعد الحمض النووي كناقل للمعلومات لتعزيز الأمن. يندمج علم إخفاء المعلومات والتشفير المستند إلى الحمض النووي بين السمات البيولوجية والتقنيات التقليدية من أجل تحقيق خوارزمية مؤمنة جيداً تستغلها. لذلك، توفر تسلسلات الحمض النووي قدرة عالية على البيانات بما في ذلك الحفاظ على الخصائص الكيميائية والبيولوجية لتسلسل الحمض النووي.

الكلمات المفتاحية

الحمض النووي، التشفير بالحمض النووي، إخفاء الحمض النووي، عملية الترميز، السبب، إخفاء البيانات



tactics based on DNA. This has spurred the development of a brand-new study area: DNA computing [8, 10].

Some of the strategies discussed in the following paragraphs are relevant to the current investigation and are used to secure data transmitted via communication channels:

- In 2010 (H. J. Shiu et) presented an approach comprising three algorithms based on DNA sequence characteristics detailed in this study. DNA sequences have certain interesting features that can be exploited to conceal information. Options include the Insertion Method, the Complementary Pair Method, and the Substitution Method. A reference DNA sequence S is chosen for each procedure, and the secret message M is placed into it, resulting in S' . Then, the sender sends the receiver S' , from which the latter may deduce and extract the message M . In addition, proof of the sturdiness and deeply rooted capacity analysis of the three proposed approaches has been provided. Extensive theoretical and empirical analysis demonstrates that the suggested methods significantly outperform the state-of-the-art alternatives [11].
- In 2011(Abbasy et al) suggested a strategy for facilitating secure resource sharing in cloud computing scenarios. The recommended approach hides information by exploiting DNA sequences. The method consists of two phases. In the first phase, the binary information is converted into a DNA sequence by applying the pairing rules to a predetermined DNA sequence. This phase, in addition to changing the data, compounds the complexity by applying the complementing rules and then indexing the jumbled sequence, which is a problem in itself. Second, using a process that is the exact opposite of the first, we will extract the secret information from the DNA sequence [12].
- System employing reversible contrast mapping was proposed in 2011 by (Hayam Mousa). The system combines two words from the sequence with the reversible contrast mapping to achieve reversibility. The method was implemented and evaluated on several different DNA sequences. The method's optimal performance was achieved when $|w|=4$. Word lengths still have the value of $|w|$ even though the sequence length has changed. And this is true regardless of the length of the series. The proposed approach can not only hide information in the DNA sequence, but also restore the original DNA sequence from the hidden findings without altering the DNA in any way. Reversible information concealment is a revolutionary method



That employing DNA in cryptography was first described in 2020 by (Baraa Tareq Hammad, et al.). Its purpose was to enhance other traditional cryptographic techniques. DNA sequences can be used to encrypt data. Two examples are shown here that investigate the use of DNA codes in tandem with traditional cryptographic techniques, each of which makes use of a distinct DNA property. The first instance employed DNA encoding using one-time pad (OTP) techniques, a kind of symmetric encryption. In the second scenario, asymmetric cryptography was investigated by adding DNA codes to the RSA method. DNA encoding was analyzed for its effectiveness in OTP, RSA, and other algorithms. DNA encoding was analyzed for its effectiveness in OTP, RSA, and other algorithms. The combined processing time of the RSA method and DNA encoding was greater, as seen by the results. To address this problem, we activated the GZIP compression method to lessen redundant data [18].

- In 2020(B. Adithya, & G. Santhi) proposed work that show how to preserve essential information while functioning in a hostile environment by combining DNA computing with Morse code encoding. Plain text is converted into DNA sequences for further analysis using an encoding table. All encoding stages for plaintext to DNA sequence conversion must be completed before initiating the encryption process. The mRNA sequence is translated into the tRNA sequence by swapping each DNA from the letter set with its DNA letter set counterpart. Using a reference DNA sequence of the same length as the ciphertext, the Study method demonstrates its superiority over the obfuscation capabilities of both the AES and RSA ciphers (181,432bp). Because of the need for extra padding bits in the encryption computation, the AES cipher can cover up to 47.02kb of reference DNA sequence, whereas the RSA cipher can cover up to 42.08kb [19].

It was planned in 2021 (Amany E. El-deed et al. The suggested data concealing approach allows for the decoding of information that has been encoded using a binary coding scheme and concealed inside a DNA sequence. In contrast to the Least Significant Base method of DNA substitution that has been discussed before. The suggested change is based on a simple concept that has never been executed before, to the best of their knowledge. As a result of this discovery, scientists discovered that DNA amino acids can be divided into groups, with each DNA codon in one of the groups being able to encode two bits of hidden information rather than just one, as the Least Significant Base approach recommends. The suggested strategy, like the Least Significant Base method, is blind, retains the DNA's normal biological structure in the false DNA sequence, and, unlike the Least Significant Base method, does not give any expansion in the DNA sequence [20].



- In 2021 (Sajib Biswas & Md. Monowar Hossain) proposed a strategy that encrypted DNA sequences. The suggested approach divided into two major steps. In the first step, they used (10 x 10) playfair cipher plus a modified Caesar cipher to encode the message into the DNA. Next, in the second step, they alter the sequence by using a general complementary base substitution approach, after which the message and cover DNA are randomly embedded and covered. After that, they insert the key into the resulting DNA at random using a random number generator. Since the recipient doesn't need to know anything going in, the extraction process may be carried out in complete darkness. Therefore, it offers us greater safety than alternative methods of secrecy [21].

- In 2021 (Shah Haris Nabi, et al) proposed technique based on DNA sequences. A DNA database is used to retrieve the reference DNA sequence. There are around 1.63 10⁸ DNA sequences that are publicly available. The proliferation of DNA sequences makes it harder for an adversary to learn whether or not a hidden message is there. Two algorithms for hiding information have been suggested. The first employs a substitution table and a reference sequence to convert top-secret information into Stego-DNA. The second strategy involves using DNA insertion to hide information. The suggested technique has the potential to improve the security of data storage, transmission, and encryption. The thesis suggests using chaos theory, cryptography, and genetic coding to make a foolproof security system. Both a 1D-2logistic map and a 4D hyperchaotic map were used in the suggested approach [22].

- In 2021(Paspula Ravinder, et al), a two-stage procedure is presented whereby the first stage involves the generation of a DNA encoding table, string matrix, and DNA digital encoding to be used in the generation of an intermediate cipher text. In the next phase, the intermediate cipher will be translated into a human mad DNA sequence and sent on to the intended recipient. With the use of an MD12-based message digest, the suggested solution additionally guarantees the data's authenticity. The approach protects the privacy, integrity, and confidentiality of information transmitted across an unsecured route of communication [23].



Table 1: The comparisons between previous works .

References	Year	Proposed work	Solved problem
[11]	2010	The three strategies offered are insertion, complementary pair, and substitution.	“Capacity, payload, and the number of bits hidden per character” (bpn).
[12]	2011	The suggested technique uses binary encoding and complementary-pair principles, both of which were implemented in the cloud.	increase confidentiality and complexity.
[13]	2011	developed a method of concealing DNA sequence information via reversible contrast mapping that can be reversed.	“The noise versus the amount of hidden data.”
[14]	2015	The proposed First DNA reference sequence is utilized to encrypt the secret message. The second sequence conceals an encrypted message. Instead of 8-bit ASCII, DNA coding encodes plaintext.	modification rate.
[15]	2015	An innovative method encodes confidential data into an audio file using an audio stenographic medium. Combining encryption with steganography improves communication.	distortion, suspicion in the mind of attackers.
[16]	2015	compared DNA-based playfair, vigenere, RSA, and AES ciphers with DNA concealing.	hiding capacity and data size.
[17]	2018	Multiple difference expansions are proposed for reversible DNA data concealing. DNA data concealing should consider string structure.	bpn of watermark, respectively, Data storage.
[18]	2020	By merging DNA codes with traditional encryption techniques, two situations were explored. First example employed DNA coding using OTP (one-time pad) techniques. In the second scenario, DNA codes were added to RSA for asymmetric cryptography.	data redundancy.
[19]	2020	“Proposed a strategy of DNA computing color code cryptography to ensure data protection from the eavesdropper. The text message in format is encrypted”	equilibrium, and energy efficiency.



[20]	2021	Substitution-based data concealing in DNA sequences was proposed. Each DNA codon in one of the groups can encode two bits.	preserves the DNA original biological structure.
[21]	2021	suggested approach operates in two primary steps. First, we encrypt the message in DNA using 10 10 Playfair. In the second stage, they modify the sequence via complementary base substitution and insert the message, and cover DNA.	Sending a keyword before the encrypted DNA Sequences.
[22]	2021	explained reversible data concealment and encryption. First, a message is hidden through DNA substitution. In data insertion, a “private key is utilized to pick a DNA sequence from the database to implant the secret message”.	storage capacity, large-scale computation.
[23]	2021	First-round intermediate encryption text is created using a DNA encoding table. “In the second round, the intermediate cipher is converted into human mad DNA and sent to the recipient”.	huge amount of data from unauthorized persons.

Discussion

Our freshly discovered DNA steganography method's major feature is its ability to hide information amid naturally occurring SNPs. We have many SNPs and many regions with high concentrations of SNPs. However, this strategy cannot be used on well-established model species with well accepted DNA sequences and few SNPs. However, other species with a high number of SNPs might also be employed to transmit covert messages. The number of SNPs needed to conceal the data was as low as 70. SNPs are more frequently needed as messages get longer.

DNA barcoding is one example of the many proposed uses for DNA steganography. There is a common misconception that DNA barcodes are intrinsic to the genome and will thus be rendered useless if the genome is altered in any way. With its block sum check-based error checking algorithm, the DNA steganography technology utilized in this work shows promise as a potential replacement for existing DNA



barcodes. In addition, "watermark" may be used to trace the origin of freshly generated cells back to a certain firm or researcher. Therefore, DNA steganography might be utilized to safeguard the exclusive nature of GMOs. Regular production of modified cells is increasing as synthetic biology and metabolic engineering technologies develop. The biotech sector relies on the ability to decipher the ownership information encoded in a cell's DNA. The data will be cloaked in highly variable SNP regions, making detection by standard techniques, such next-generation sequencing, impossible (NGS). Because the approach may pick up on mistakes, spotting alterations to the coded message is a breeze. This suggests that the suggested DNA steganography technique may be a practical means of shielding complex organisms.

Conclusion

DNA-based methods of information concealment enhance security by shielding stored data from prying eyes. DNA-based security is an emerging information security technique with several advantageous features, including large capacity, high randomness, and low modification rate. Therefore Information can be encoded in a non-detectable fashion in biological carriers for a variety of purposes, including but not limited to verifying the authenticity of DNA sequences, annotating sequences for streamlined cataloging and research, and covert data transmission. DNA has emerged as a promising new data carrier in recent years because of its efficiency and reliability. DNA's bio-molecular computing skills can be used for cryptography and stenography to create low-cracking-probability, high-capacity secured algorithms. In this study, the authors evaluate and contrast a number of recently proposed DNA-based steganography algorithms that focus on a few central, defining characteristics. The primary goal of this comparison was to assist researchers to address existing algorithms' limitations to advance the state-of-the-art in good and enhanced secured DNA steganography approaches.

Conflict of interests.

There are non-conflicts of interest.

References

- [1] Singh, G. (2013). A study of encryption algorithms (RSA, DES, 3DES and AES) for information security. *International Journal of Computer Applications*, 67(19).
- [2] Subhedar, M. S., & Mankar, V. H. (2014). Current status and key issues in image steganography: A survey. *Computer science review*, 13, 95-113.
- [3] Hamed, G., Marey, M., El-Sayed, S. A., & Tolba, M. F. (2016, December). Comparative study for various DNA based steganography techniques with the essential conclusions about the future research. In *2016 11th International Conference on Computer Engineering & Systems (ICCES)* (pp. 220-225). IEEE.



- [4] Amin, M. M., Salleh, M., Ibrahim, S., Katmin, M. R., & Shamsuddin, M. Z. I. (2003, January). Information hiding using steganography. In *4th National Conference of Telecommunication Technology, 2003. NCTT 2003 Proceedings.* (pp. 21-25). IEEE.
- [5] Al-Mohammad, A. (2010). *Steganography-based secret and reliable communications: Improving steganographic capacity and imperceptibility* (Doctoral dissertation, Brunel University, School of Information Systems, Computing and Mathematics Theses).
- [6] Hamed, G., Marey, M., El-Sayed, S. A., & Tolba, M. F. (2015, November). Hybrid technique for steganography-based on DNA with n-bits binary coding rule. In *2015 7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)* (pp. 95-102). IEEE.
- [7] Anam, B., Sakib, K., Hossain, M., & Dahal, K. (2010). Review on the Advancements of DNA Cryptography. *arXiv preprint arXiv:1010.0186*.
- [8] - Hamed, G., Marey, M., El-Sayed, S., & Tolba, F. (2016). DNA based steganography: survey and analysis for parameters optimization. In *Applications of intelligent optimization in biology and medicine* (pp. 47-89). Springer, Cham.
- [9] Das, S., Das, S., Bandyopadhyay, B., & Sanyal, S. (2011). Steganography and Steganalysis: different approaches. *arXiv preprint arXiv:1111.3758*.
- [10] Peterson, I. (2001). Hiding in DNA. *Proceedings of Muse*, 22.
- [11] Shiu, H. J., Ng, K. L., Fang, J. F., Lee, R. C., & Huang, C. H. (2010). Data hiding methods based upon DNA sequences. *Information Sciences*, 180(11), 2196-2208.
- [12] Abbasy, M. R., & Shanmugam, B. (2011, July). Enabling data hiding for resource sharing in cloud computing environments based on DNA sequences. In *2011 IEEE World Congress on Services* (pp. 385-390). IEEE.
- [13] Mousa, H., Moustafa, K., Abdel-Wahed, W., & Hadhoud, M. M. (2011). Data hiding is based on contrast mapping using a DNA medium. *Int. Arab J. Inf. Technol.*, 8(2), 147-154. [14]- Ibrahim, F. E., Abdalkader, H. M., & Moussa, M. I. (2015). Enhancing the security of data hiding using double DNA sequences. In *Industry-Academia Collaboration Conference (IAC)* (pp. 6-8).
- [15] Sahana, S., Dey, G., Ganguly, M., Paul, P., & Paul, S. Adaptive Steganography Based Enhanced Cipher Hiding Technique for Secure Data Transfer.
- [16]- Marwan, S., Shawish, A., & Nagaty, K. (2016). DNA-based cryptographic methods for data hiding in DNA media. *Biosystems*, 150, 110-118.
- [17] Lee, S. H., Lee, E. J., Hwang, W. J., & Kwon, K. R. (2018). Reversible DNA data hiding using multiple difference expansions for DNA authentication and storage. *Multimedia Tools and Applications*, 77(15), 19499-19526.
- [18] Hammad, B. T., Sagheer, A. M., Ahmed, I. T., & Jamil, N. (2020). A comparative review on symmetric and asymmetric DNA-based cryptography. *Bulletin of Electrical Engineering and Informatics*, 9(6), 2484-2491.
- [19] Adithya, B., & Santhi, G. (2021). Deoxyribonucleic Acid (DNA) Computing using Two-by-Six Complementary and Color Code Cipher. *Bulletin of Computer Science and Electrical Engineering*, 2(1), 38-45.
- [20]- (A Substitution-Based Method for Data Hiding in DNA Sequences) A El-Deeb, A Elsisy, A Youssef - IJCI. International Journal , 2021 - journals.ekb.eg.
- [21] Biswas, S., & Hossain, M. (2021). Fully Blind Data Hiding by Embedding Within DNA Sequences Using Various Ciphering and Generic Complimentary Base Substitutions. In *Proceedings of*



- International Joint Conference on Advances in Computational Intelligence* (pp. 1-13). Springer, Singapore.
- [22] Nabi, S. H., Sarosh, P., Parah, S. A., & Mohiuddin Bhat, G. (2021). Information Embedding Using DNA Sequences for Covert Communication. In *Multimedia Security* (pp. 111-129). Springer, Singapore.
- [23] Ravinder, P., Yadav, M. G., Rashmi, M. K., & Srividhya, M. C. (2021). Information Hiding through DNA Sequence Technology. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal/ NVEO*, 5678-5687.