Biosafety and Health 6 (2024) 251–256

Contents lists available at ScienceDirect

Biosafety and Health

journal homepage: www.elsevier.com/locate/bsheal

Review Article

Cyberbiosecurity: Advancements in DNA-based information security

Tuoyu Liu [a,b], Sijie Zhou [c,d], Tao Wang [b], Yue Teng [a,*,1]

[a] State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing 100071, China
[b] School of Life Sciences, Tianjin University, Tianjin 300072, China
[c] Frontiers Science Center for Synthetic Biology and Key Laboratory of Systems Bioengineering (Ministry of Education), School of Chemical Engineering and Technology, Tianjin University, Tianjin 300072, China
[d] Frontiers Research Institute for Synthetic Biology, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Article history:
Received 31 January 2024
Revised 14 June 2024
Accepted 19 June 2024
Available online 21 June 2024

Keywords:
Synthetic biology
Cyberbiosecurity
Information security
Deoxyribonucleic acid (DNA) storage
Deoxyribonucleic acid (DNA) sequencing

ABSTRACT

Synthetic biology is a crucial component of the "cyber-biological revolution" in this new industrial revolution. Owing to breakthroughs in synthetic biology, deoxyribonucleic acid (DNA), the storehouse of hereditary material in biological systems, can now be used as a medium for storage (synthesis) and reading (sequencing) of information. However, integrating synthetic biology with computerization has also caused cyberbiosecurity concerns, encompassing biosecurity and information security issues. Malicious codes intended to attack computer systems can be stored as artificially synthesized DNA fragments, which can be released during DNA sequencing and decoding and attack computer and network systems. As these cyberbiosecurity threats become increasingly realistic, spreading awareness and information about how they can be prevented and controlled is crucial. This review aims to address this need by offering crucial theoretical backing for cyberbiosecurity research and raising awareness of risk mitigation and control measures in information security, biosecurity, and national security.

1. Introduction

Synthetic biology is an emerging discipline of the 21st century that has led to remarkable achievements in bioenergy, biomaterials, and biomedical engineering and the exploration of the essence of life. One significant advancement in chemical and molecular biology technologies is the artificial synthesis of whole genomes of simple life forms, such as bacteria and viruses. D.G. Gibson et al. [1] reported the world's first bacterial genome entirely chemically synthesized and assembled de novo in the United States [1]. These achievements can be attributed to the rapid development of and advancements in deoxyribonucleic acid (DNA) synthesis and sequencing technologies. Furthermore, the rapid evolution of next-generation sequencing (NGS) technology, used to determine the sequence of bases in DNA fragments, has led to a continuous fall in the cost of DNA sequencing [2]. For example, the development of the Illumina sequencing platform has reduced the cost of sequencing the human genome from roughly US$100,000 to US$1,000 [3]. These breakthroughs have revolutionized the development of new interdisciplinary fields, like DNA computing and storage, and transformed synthetic biology technologies. Notably, the shortening of DNA synthesis time, improvement of DNA sequencing accuracy, and continuous decline of DNA synthesis and sequencing costs have been crucial driving factors in the development of DNA-based data storage technology, which has become increasingly feasible and is widely applied. Creating a DNA storage system involves writing and storing encoded data as artificially synthesized DNA molecules; data stored in this way can be retrieved and read by DNA sequencing [4]. Compared to traditional information storage methods, DNA-based data storage offers higher storage density, longer data retention, and faster data copy [5,6].

As with many emerging technologies, synthetic biology poses a dual-use dilemma [7]. While it realizes the artificial creation and transformation of life forms through gene synthesis and gene editing, the swift progress in synthetic biology raises the potential for misuse, abuse, and concerns regarding biosecurity in the scientific community [8,9]. With the decreasing cost and simplification of gene synthesis technology, biohackers will efficiently conduct the artificial synthesis of a new virus or the de novo creation of a virus using natural viruses as a reference [10,11]. Therefore, combining synthetic biology, characterized by highly compatible technical standards and open material data resources, with highly pathogenic microbes may significantly threaten biosecurity. The US National Academy of Sciences examined

* Corresponding author: State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, 20 Dong-Da Street, Fengtai District, Beijing 100071, China.
E-mail address: yueteng@sklpb.org (Y. Teng).

[1] Given his role as the Guest Editor, Yue Teng had no involvement in the peer-review of this article and had no access to information regarding its peer-review. Full responsibility for the editorial process for this article was delegated to Editor Yong Zhang.

https://doi.org/10.1016/j.bsheal.2024.06.002
2590-0536/

the biosecurity threats connected with synthetic biology in its 2018 report, *Biodefense in the Age of Synthetic Biology* [12]. China also passed the *Biosecurity Law* in 2021 in response to the biosecurity issue, stipulating the strengthening of biotechnology research standards and the prevention of threats from bioterrorism and bioweapons.

With the continuous cross-integration of technologies, such as gene synthesis, gene sequencing, bioinformatics, and computer information technology, artificially synthesized DNA molecules present cyberbiosecurity issues during the highly integrated DNA synthesis, sequencing, and computer analysis processes. Specifically, artificially synthesized DNA can be programmed to carry malicious programs that exploit system vulnerabilities. While reading these DNA fragments, codes are released to execute attacks on the computers and networks involved in the related processes [13]. These attacks treat DNA as digital information, and the computer systems and networks involved in the synthesis and sequencing processes are targeted similarly to conventional computer information security breaches. Remaining vigilant about DNA synthesis and storage technologies that might deviate from the set scientific research objectives is essential to prevent their misuse. Additionally, defenses within DNA synthesis and sequencing technologies must be bolstered to prepare for forthcoming challenges in this emerging security domain [14].

While research articles and corresponding solutions and defense strategies regarding cyberbioattacks are already available, they are still relatively rudimentary compared to traditional computer network security measures [15–17]. Current studies predominantly focus on the potential applications of synthetic biology and DNA storage systems. Yet, in-depth and comprehensive discussions on interdisciplinary consequences are lacking, particularly at the intersection of cyberbiosecurity threats and traditional biosecurity issues. We must remain alert to prevent and prepare for the possibility of future, more subtle and advanced attacks. This review aims to provide a theoretical foundation for research into cyberbiosecurity and offer strategic insights for addressing threats within the realms of information security, biosecurity, and national defense. It starts with an introduction to the technical principles of DNA synthesis and sequencing, followed by an outline of the threats to cyberbiosecurity, a discussion on defense strategies against these threats, and the anticipated challenges in cyberbiology.

2. Basic principles of DNA synthesis and sequencing

Synthetic biology is the study of artificial biological systems; it draws upon the genetic engineering and engineering methodologies of systems biology. It entails applying engineering concepts and techniques to biotechnological domains, including genetic and cell engineering.

Genetic material can be created directly from information and simple chemical components thanks to DNA synthesis methods. Presently, the phosphoramidite trimester method is the prevalent approach for DNA synthesis, prized for its speed and efficiency in the chemical creation of DNA [18]. Y. Shao et al. [19] synthesized a *Saccharomyces cerevisiae* strain with only one chromosome and completed the recombination of its 11.8-Mb genome in 2018 [19]. By successfully integrating 6.5 synthetic yeast chromosomes into a single yeast cell, Y. Zhao et al. [20] made great strides in synthetic biology and demonstrated an advanced approach for building and functionally integrating synthetic chromosomes into living organisms [20]. The reaction involves linking a single base to the DNA fragment on a solid-phase carrier via deblocking, activation, coupling, capping, and oxidation. Crude DNA fragments can be obtained by repeating these steps. These fragments are then processed by cutting, removing protective groups, and purifying the fragments to get the target DNA fragments. Usually, a plasmid system is used to build and insert the synthesized DNA fragments. The resulting plasmid is very stable and can be transferred into living cells, where the fragments are replicated [21]. Advances in DNA synthesis technology have catalyzed innovations and breakthroughs in many research fields, as the technology allows for de novo gene design and construction and the integration and assembly of gene circuits without the need for existing DNA templates [22].

Unlike DNA synthesis, which involves a de novo gene construction, DNA sequencing aims to decipher the inherent order of bases in a DNA specimen. DNA sequencing technology comprises three categories: first-generation Sanger sequencing, second-generation high-throughput sequencing (or NGS) represented by the Illumina sequencing platform, and third-generation single-molecule sequencing [23,24].

The widely used second-generation sequencing technology can currently read many relatively short sequences (fragments) in parallel to sequence DNA samples. The sequencing steps are as follows: (1) The long sequences are cut into fragments and amplified by polymerase chain reaction (PCR); (2) A DNA adapter sequence (a short, known piece of DNA used to provide a binding site for sequencing primers) is attached at the termini of the amplified DNA fragments; (3) The bi-stranded DNA gets disassembled into individual strands before being introduced into a glass flow cell. At this point, the adapter sequence interacts with matching fragments on the cell surface, undergoing local replication and generating the same DNA clusters; (4) Subsequently, fluorescence-labeled nucleotides are inserted at the ends of the individual DNA strands in these clusters, and the fluorescence is observed to ascertain the DNA sequence in each group, and (5) The base sequences identified by fluorescence are stored in the FASTQ text format. Although NGS sequencing is cost-effective, swift, and accurate, it has the limitation that it is a short-read sequencing method in which fragment lengths are strictly limited. However, third-generation sequencing technologies, such as Oxford Nanopore sequencing, which can support long-fragment sequencing, have been developed and can overcome this limitation [25]. Single-molecule sequencing can sequence a considerably longer fragment and is set to change sequencing data structure fundamentally and enhance future sequencing capability.

DNA synthesis and sequencing technologies that provide higher accuracy, the ability to synthesize or read longer DNA fragments, and lower operation costs are increasingly becoming available and paving the way to the wide application of DNA storage technology. In 2019, Microsoft Research and the University of Washington in the US realized fully automated DNA data storage and extraction for the first time [26]. In DNA storage technology, the DNA fragments used to store data files must first be digitally encoded; digitized information is converted into the four DNA bases, with their sequence corresponding to the digitized information in the digital file. When encoding the information, the physical constraints of DNA and the maximum acceptable length of a synthetic DNA strand must be considered. DNA synthesis technology can synthesize the encoded sequence into a DNA strand to store information. This artificial DNA is inserted into living cells (*in vivo*) or stored in a DNA pool (*in vitro*). For *in vitro* storage, a large DNA database can be constructed. Array-based DNA synthesis can be employed when many different DNA sequences need to be synthesized, as this technology can synthesize several unique sequences in parallel [27]. The synthetic DNA is sequenced using a computer for data retrieval, and the obtained sequences are decoded. This technology's advantages are the continuous optimization of coding error correction capabilities, gradual cost reduction, and the possibility of integrating artificial intelligence technologies. In the future, DNA storage will likely profoundly impact the global data field.

3. Emerging cyberbiosecurity threats

Synthetic biology has intersected and merged with various fields, including genetics, molecular biology, systems biology, bioinformat-

ics, genetic engineering, and metabolic engineering, thanks to the advancement of DNA synthesis and sequencing technologies. This convergence has produced many noteworthy breakthroughs [28–33]. Processing and analyzing enormous amounts of biological data, which usually require computers to handle large amounts of computation, is essential for these achievements. Typically, the sequencer is a computer or is connected to a computer to complete its sequencing tasks. Additionally, DNA sequencing files require various computer software for processing and executing large-scale analysis tasks. Synthetic DNA fragments could be deployed as weapons to target related computer programs during the sequence analysis, creating a new cyber-biological threat as a result of the junction of DNA storage and computer technology. This method of attack leverages the inherent property of DNA molecules as an information medium. DNA can encode and store malicious programs, which can remotely compromise related systems and networks when the "contaminated" DNA is "input" into computer systems [13].

DNA creation, manipulation, processing, and analysis steps are prone to attacks similar to those in conventional information security. This emerging attack method is difficult to detect during the attack process, and this problem is compounded by lacking defensive measures [34]. P. Ney et al. [14] from the University of Washington in the US conducted a study in which they attacked the sequencing process using synthetic DNA containing a malicious program [14]. First, the author modified a DNA-processing program by intentionally introducing a software vulnerability. This vulnerability might then be used to simulate real vulnerability risks in the software by executing codes stored in DNA. Next, a DNA sequence containing the malicious program was synthesized and sequenced using NGS; after sequencing, a FASTQ file containing the malicious program was generated [35]. Short DNA fragments were analyzed by the malicious application, which was launched when the file was read and the sequences were processed. The results indicated that manipulating and processing a DNA sequencing file containing a malicious program could seriously threaten the computer analysis system. The specific attack procedure used in the previous study is presented in Fig. 1.

P. Ney et al. also analyzed the security of the NGS sequencing file processing procedure. This procedure includes preprocessing, alignment, de novo assembly, post-alignment processing, ribonucleic acid (RNA) sequence analysis, and binding site analysis. The study found multiple insecure function calls and buffer overflow risks (such as the C function of "strcat strcpy sprint") in programs analyzing NGS sequencing results that could be targeted by attacks [36]. Additionally, they examined three well-used software programs to demonstrate that overflows could be caused by the static buffers in these systems. (1) The Fastx toolkit generates summary statistics for FASTQ files. They found that the buffer overflow check setting does not align with the actual buffer size, which may render it ineffective against overflows. (2) Samtools, a suite of tools for handling Sam and Bam files. The same buffer is used for input and compressed headers, and buffer boundary issues might arise if the input headers are incorrectly formatted. (3) SOAPdenovo2, a de novo genome assembly software, where the buffers lack boundary checks. In cybersecurity, buffer overflow vulnerabilities pose significant security risks and are prime attack targets. The frequent unsecured function calls and lack of buffer checks detected in the NGS processing point to the current software programs' deficiency in security standards. Moreover, buffer overflow vulnerabilities are widespread in NGS programs. As a result of these security gaps, using malicious code stored in DNA to exploit the vulnerabilities above is highly feasible and could have severe consequences.

Some companies provide technical support for DNA synthesis and sequencing analysis through remote and cloud services. They can offer integrated services for DNA synthesis and sequencing through mutual cooperation. In 2020, the DNA Data Storage Alliance was established by 15 companies, including sequencing firms represented by Illumina, DNA synthesis firms represented by Twist Bioscience, and data storage companies represented by Microsoft Research and Western Digital. They intended to create and promote an automated storage ecosystem based on artificially synthesized DNA as a data storage medium. However, these systematic integrations inadvertently increased the risk of cyberbiosecurity attacks. Furthermore, inherent risks to DNA data sharing were already present. For example, sequencing data generated by various research teams are often analyzed and deposited in biological databases and shared via email. These shared data can be a potential means of attack, even without synthesizing DNA sequences [37]. Thus, along with the increasing application of such integrated technologies, remaining vigilant about potential threats and creating strategies to prevent and control these risks is crucial.

4. Directions for cyberbiosecurity defense

Biocomputer attacks are similar to conventional cyberattacks but are biologically constrained due to using DNA as the storage medium. Therefore, the inherent characteristics of defensive strategies used in cybersecurity can be adopted to develop targeted defense methods for such biocomputer attacks. The following four measures can be employed to mitigate or even prevent these attacks, according to the defense mechanisms commonly utilized in computer network security. First, users should do routine security audits and software updates, and bioinformatics software creation and use should adhere to secure software guidelines. Second, software packages should be signed to confirm their genuineness. When sharing files, the sequencing document could have a digital verification tag or encryption added by the originating research group [38]. It could then be uploaded to the database or distributed to other research units. Third, the security of DNA sequencing samples must be guaranteed by close monitoring from collection to sequencing. Fourth, checking whether the synthetic sequence can be decoded into malicious codes during and before synthesis is also a potential solution. DNA synthesis programs have a strict review system to ensure that organisms like pathogens, toxins, and other harmful products are not produced through DNA synthesis and splicing. Malicious codes within DNA can be detected using similar methods, but detecting and identifying malicious codes in DNA sequences during the synthesis phase is not simple [39].

Biocomputer attacks are a threat; however, they are limited by the inherent properties of DNA synthesis and reading. Because attackers usually need to add additional content to the DNA sequence to exploit system vulnerabilities, these biological constraints can be leveraged against them. For instance, any vulnerabilities rooted in DNA must be concretely represented within the DNA itself; the design of DNA strands must adhere to the relevant criteria of synthetic DNA to enable the successful synthesis of harmful sequences. These specifications include limiting the content of the nucleotide pairs GC and avoiding excessively long continuous stretches of AT or GC pairs. Typically, a high GC content can result in DNA fragments not effectively unwinding or separating even at lower temperatures, which may lead to secondary structures. Meanwhile, long consecutive stretches of AT or GC can cause errors during synthesis, as such segments are more prone to mistakes during synthesis and may pose challenges for subsequent DNA sequencing. Furthermore, preventing single strands from forming secondary structures on their own during synthesis is essential. Additionally, whether the information encoded in DNA can be accurately deciphered is directly impacted by the randomness and errors in the sequencing process. Attackers typically use symmetrical strands on both ends, resynchronized instruction sequences, and error correction codes to overcome these inherent restrictions. However, these operations limit coding complexity and flexibility, resulting in synthetic DNA strands with similar characteristics. Nevertheless, as second- and third-generation sequencing advances and gains traction, computers used for sequence processing and analysis will be more sus-

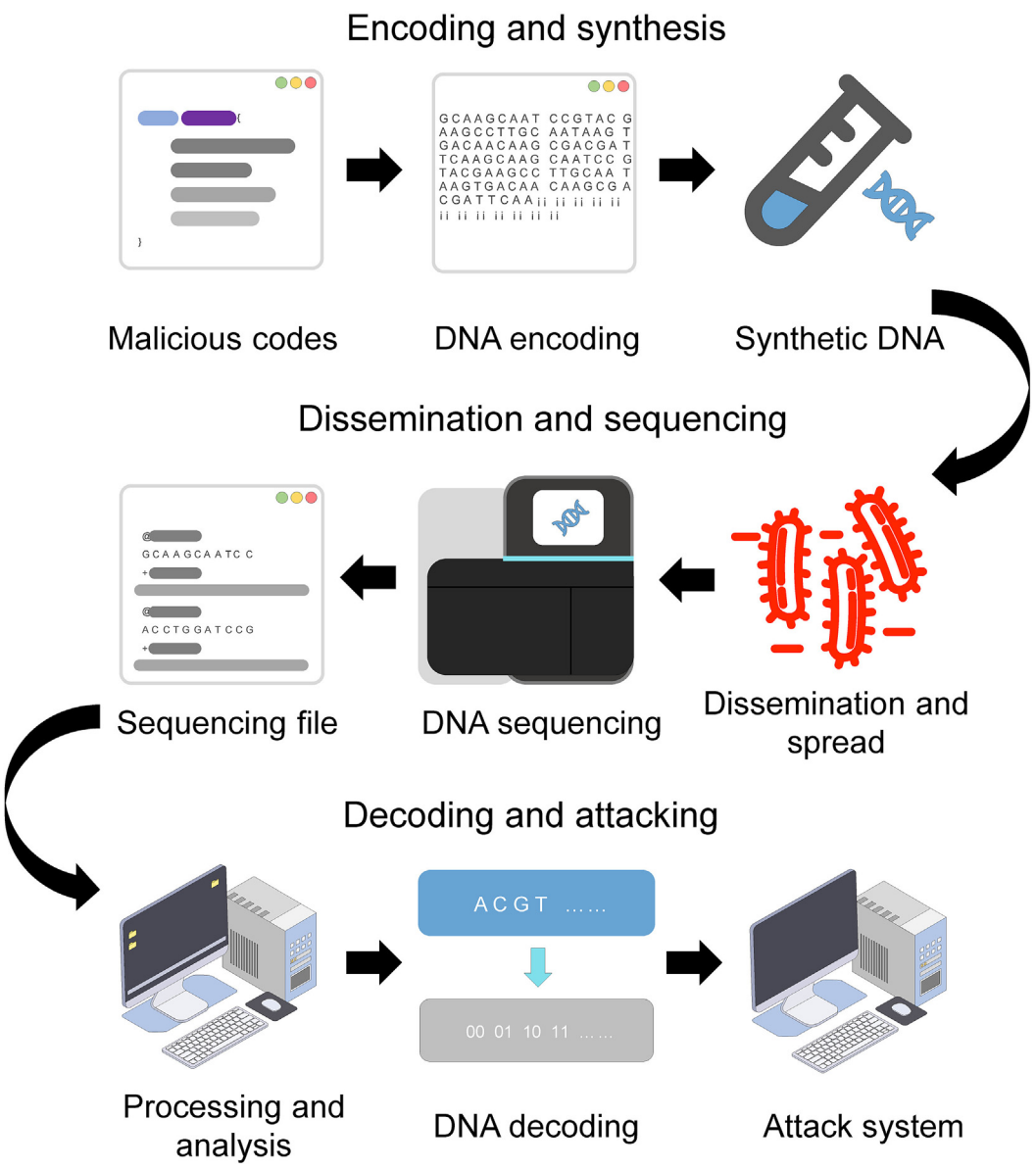

Fig. 1. Flowchart depicting a cyberbioattack. Malicious codes can be encoded in the deoxyribonucleic acid (DNA) fragment and compromise computer equipment when DNA is synthesized and sequenced.

ceptible to attacks. Thus, sequencing output data and analysis software programs must be more carefully supervised. Effective detection and identification methods should be used to screen the output data before analyzing sequencing output files to ensure their safety.

5. Summary and outlook

DNA sequencing has become widely used as a result of technological advancements, especially the ongoing development of high-precision and long-read sequencing. Concurrently, DNA synthesis technology has further facilitated the practical application of DNA storage. However, these developments have also given attackers additional chances to target cyber-bio systems; attacks using synthetic DNA as a vector are particularly expected to rise. On the one hand, DNA synthesis and sequencing technologies might be utilized as tools for attack; on the other hand, they could become targets of attacks.

Many countries have incomplete prevention capabilities and detection mechanisms against biosecurity threats. For example, the *Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA* released by the US Department of Health and Human Services aims to reduce the risks of nucleic acid synthesis technology being used maliciously through screening technologies [40]. Regarding the risks and protection of biological data, the *European Union's General Data Protection Regulation* provides explicit norms for genetic data, emphasizing the need for the special protection of individual genetic data [41,42]. In an effort to prevent foreign adversaries from accessing American genetic data, the US House Special Committee on Strategic Competition between the US and China proposed the *BIOSECURE Act*, which prohibits US government agencies from purchasing or using biotechnology equipment or services from these businesses. In light of the possible advantages and disadvantages of employing government-owned biological databases for biomedical research, a Biden executive order underlined the significance of understanding and leveraging arti-

ficial intelligence to prevent biological dangers. These policies and regulatory measures aim to ensure that the development of biotechnology is safe and ethical. Yet, awareness about adequate defense measures against emerging cyber-biological cross-attacks is lacking. Therefore, compromising information and national security through attacks on computers and networks remains at risk. Research on cybersecurity based on DNA molecules is still in its initial stages in China, and systematic studies on the theoretical and methodological basis are lacking. Theoretical work on the viability of exploiting synthetic DNA to breach computer systems and security studies of the DNA sequencing procedure can provide a basis for developing countermeasures for cyberbiosecurity risks. This requires a focused investigation of potential defense mechanisms that may influence the direction of future bioinformatics design and development. Notably, studies on information security based on DNA molecules will significantly enhance the defensive capabilities of nations regarding information security, biosecurity, and national security.

Acknowledgements

We thank our colleagues at the State Key Laboratory of Pathogen and Biosecurity for the insightful discussions, and our colleagues at the Beijing Institute of Microbiology and Epidemiology for their technical assistance.

Conflcit of interest statement

The authors declare that there are no conflicts of interest.

Author Contributions

Tuoyu Liu: Literature Search, Data analysis, Writing-original draft. **Sijie Zhou:** Writing-review & editing. **Tao Wang:** Writing-review & editing. **Yue Teng:** Conceptualization, Supervision, Writing-review & editing.

References

[1] D.G. Gibson, G.A. Benders, C. Andrews-Pfannkoch, E.A. Denisova, H. Baden-Tillson, J. Zaveri, T.B. Stockwell, A. Brownley, D.W. Thomas, M.A. Algire, et al., Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome, Science (New York, N.Y.) 319 (5867) (2008) 1215–1220, https://doi.org/10.1126/science.1151721.

[2] K.H. Redford, W. Adams, R. Carlson, G.M. Mace, B. Ceccarelli, Synthetic biology and the conservation of biodiversity, Oryx 48 (3) (2014) 330–336, https://doi.org/10.1017/S0030605314000040.

[3] K.A. Wetterstrand, DNA sequencing costs: Data from the NHGRI Genome Sequencing Program (GSP), National Human Genome Research Institute. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data, 2013 (accssed 15 January 2024).

[4] V. Zhirnov, R.M. Zadegan, G.S. Sandhu, G.M. Church, W.L. Hughes, Nucleic acid memory, Nat. Mater. 15 (4) (2016) 366–370, https://doi.org/10.1038/nmat4594.

[5] M.G. Rutten, F.W. Vaandrager, J.A. Elemans, R.J. Nolte, Encoding information into polymers, Nat. Rev. Chem. 2 (11) (2018) 365–381, https://doi.org/10.1038/s41570-018-0051-5.

[6] Y. Teng, S. Yang, J. Li, Y. Cui, R. Liu, S. Wang, Principle and progress of DNA data storage, Prog. Biochem. Biophysics 48 (05) (2021) 494–504, https://doi.org/10.16476/j.pibb.2020.0224.

[7] J.R. Clapper, Statement for the record: Worldwide threat assessment of the US intelligence community. https://www.armed-services.senate.gov/imo/media/doc/Clapper_02-09-16.pdf, 2016 (accessed 15 January 2024).

[8] T.M. Tumpey, C.F. Basler, P.V. Aguilar, H. Zeng, A. Solórzano, D.E. Swayne, N.J. Cox, J.M. Katz, J.K. Taubenberger, P. Palese, Characterization of the reconstructed 1918 Spanish influenza pandemic virus, Science 310 (5745) (2005) 77–80, https://doi.org/10.1126/science.1119392.

[9] G.K. Gronvall, Synthetic biology: biosecurity and biosafety implications, in: S. Singh, J. Kuhn (Eds.), Defense Against Biological Attacks, Springer, Cham, 2019, pp. 225–232.

[10] R.J. Jackson, A.J. Ramsay, C.D. Christensen, S. Beaton, D.F. Hall, I.A. Ramshaw, Expression of mouse interleukin-4 by a recombinant ectromelia virus suppresses cytolytic lymphocyte responses and overcomes genetic resistance to mousepox, J. Virol. 75 (3) (2001) 1205–1210, https://doi.org/10.1128/JVI.75.3.1205-1210.2001.

[11] J. Cello, A.V. Paul, E. Wimmer, Chemical synthesis of poliovirus cDNA: Generation of infectious virus in the absence of natural template, Science 297 (5583) (2002) 1016–1018, https://doi.org/10.1126/science.1072266.

[12] National Academies of Sciences, Engineering and Medicine, Biodefense in the age of synthetic biology, National Academies Press 10, Washington DC, 2018.

[13] D. Farbiash, R. Puzis, Cyberbiosecurity: DNA injection attack in synthetic biology [Preprint], arXiv (2020) 2011.14224, https://doi.org/10.48550/arXiv.2011.14224.

[14] P. Ney, K. Koscher, L. Organick, L. Ceze, T. Kohno, Computer security, privacy, and DNA sequencing: Compromising computers with synthesized DNA, privacy leaks, and more, Proceedings of the 26th USENIX Conference on Security Symposium, USENIX Association, Vancouver, BC, Canada, 2017, pp. 765–779.

[15] H. Bae, S. Min, H.-S. Choi, S. Yoon, DNA privacy: Analyzing malicious DNA sequences using deep neural networks, IEEE/ACM Trans. Comput. Biol. Bioinf. 19 (2) (2020) 888–898, https://doi.org/10.1109/TCBB.2020.3017191.

[16] M.S. Islam, S. Ivanov, H. Awan, J. Drohan, S. Balasubramaniam, L. Coffey, S. Kidambi, W. Sri-saan, Using deep learning to detect digitally encoded DNA trigger for Trojan malware in Bio-Cyber attacks, Sci. Rep. 12 (1) (2022) 9631, https://doi.org/10.1038/s41598-022-13700-5.

[17] J.H.D.B. Gervasio, H. da Costa Oliveira, A.G. da Costa Martins, J.B. Pesquero, B.M. Verona, N.N.P. Cerize, How close are we to storing data in DNA?, Trends Biotechnol 42 (2023) 156–167, https://doi.org/10.1016/j.tibtech.2023.08.001.

[18] M. Caruthers, A. Barone, S. Beaucage, D. Dodds, E. Fisher, L. McBride, M. Matteucci, Z. Stabinsky, J.-Y. Tang, Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method, Methods Enzymol 154 (1987) 287–313, https://doi.org/10.1016/0076-6879(87)54081-2.

[19] Y. Shao, N. Lu, Z. Wu, C. Cai, S. Wang, L.-L. Zhang, F. Zhou, S. Xiao, L. Liu, X. Zeng, Creating a functional single-chromosome yeast, Nature 560 (7718) (2018) 331–335, https://doi.org/10.1038/s41586-018-0382-x.

[20] Y. Zhao, C. Coelho, A.L. Hughes, L. Lazar-Stefanita, S. Yang, A.N. Brooks, R.S. Walker, W. Zhang, S. Lauer, C. Hernandez, Debugging and consolidating multiple synthetic chromosomes reveals combinatorial genetic interactions, Cell 186 (2023) 5220-5236. e16, https://doi.org/10.1016/j.cell.2023.09.025.

[21] W. Mandecki, M.A. Hayden, M.A. Shallcross, E. Stotland, A totally synthetic plasmid for general cloning, gene expression and mutagenesis in Escherichia coli, Gene 94 (1) (1990) 103–107, https://doi.org/10.1016/0378-1119(90)90474-6.

[22] S. Yang, J. Li, Y. Cui, Y. Teng, The current status and future prospects of DNA computing, Sheng Wu Gong Cheng Xue Bao 37 (4) (2021) 1120–1130, https://doi.org/10.13345/j.cjb.200408.

[23] S. Behjati, P.S. Tarpey, What is next generation sequencing?, Arch Dis. Child Educ. Pract. Ed. 98 (6) (2013) 236–238, https://doi.org/10.1136/archdischild-2013-304340.

[24] P. Du, W.A. Kibbe, S.M. Lin, lumi: a pipeline for processing Illumina microarray, Bioinformatics 24 (13) (2008) 1547–1548, https://doi.org/10.1093/bioinformatics/btn224.

[25] M. Jain, H.E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community, Genome Biol. 17 (2016) 1–11, https://doi.org/10.1186/s13059-016-1103-0.

[26] C.N. Takahashi, B.H. Nguyen, K. Strauss, L. Ceze, Demonstration of end-to-end automation of DNA data storage, Sci. Rep. 9 (1) (2019) 4998, https://doi.org/10.1038/s41598-019-41228-8.

[27] S. Kosuri, G.M. Church, Large-scale de novo DNA synthesis: technologies and applications, Nat. Methods 11 (5) (2014) 499–507, https://doi.org/10.1038/nmeth.2918.

[28] J.M. Wagner, H.S. Alper, Synthetic biology and molecular genetics in non-conventional yeasts: current tools and future advances, Fungal Genet Biol. 89 (2016) 126–136, https://doi.org/10.1016/j.fgb.2015.12.001.

[29] B.L. Adams, The next generation of synthetic biology chassis: moving synthetic biology from the laboratory to the field, ACS Synth. Biol. 5 (12) (2016) 1328–1330, https://doi.org/10.1021/acssynbio.6b00256.

[30] Y. Liu, H.-D. Shin, J. Li, L. Liu, Toward metabolic engineering in the context of system biology and synthetic biology: Advances and prospects, Appl. Microbiol. Biotechnol. 99 (2015) 1109–1118, https://doi.org/10.1007/s00253-014-6298-y.

[31] J. Nicholas, Bioinformatics for the synthetic biology of natural products: Integrating across the design–build–test cycle, Nat. Prod. Rep. 33 (8) (2016) 925–932, https://doi.org/10.1039/c6np00018e.

[32] R.M. West, G.K. Gronvall, CRISPR cautions: Biosecurity implications of gene editing, Perspect Biol Med. 63 (1) (2020) 73–92, https://doi.org/10.1353/pbm.2020.0006.

[33] R. García-Granados, J.A. Lerma-Escalera, J.R. Morones-Ramírez, Metabolic engineering and synthetic biology: Synergies, future, and challenges, Front. Bioeng. Biotechnol. 7 (2019) 36, https://doi.org/10.3389/fbioe.2019.00036.

[34] I. Fayans, Y. Motro, L. Rokach, Y. Oren, J. Moran-Gilad, Cyber security threats in the microbial genomics era: Implications for public health, Euro Surveill. 25 (6) (2020) 1900574, https://doi.org/10.2807/1560-7917.ES.2020.25.6.1900574.

[35] A. One, Smashing the stack for fun and profit, Phrack Magazine 7 (49) (1996) 14–16, https://doi.org/10.1007/978-3-319-99828-2_21.

[36] M.M. Sarnowski, D. Larson, S.M. Alnaeli, M.K. Sarrab, A study on the usage of unsafe functions in gcc compared to mobile software systems, in: 2017 IEEE International Conference on Electro Information Technology (EIT), 2017, pp. 138–142.

[37] A. Delgado, DIYbio: Making things and making futures, Futures 48 (2013) 65–73, https://doi.org/10.1016/j.futures.2013.02.004.

[38] A. Elhadad, A. Khalifa, S. Rida, DNA-based data encryption and hiding using playfair and insertion techniques, J. Commun. Comput. Eng. 2 (3) (2012) 44, https://doi.org/10.20454/jcce.2012.242.10.20454/jcce.2012.242.

[39] J. Mason, S. Small, F. Monrose, G. MacManus, English shellcode, proceedings of the 16th ACM conference on Computer and communications security, 2009, pp. 524–533.
[40] C.M. Sharkey, M. Lekveishvili, T. de la Rosa, K. Danskin, Enhancing gene synthesis security: An updated framework for synthetic nucleic acid screening and the responsible use of synthetic biological materials, Appl. Biosafety 29 (2024) 63–70, https://doi.org/10.1089/apb.2023.0036.
[41] F.D. Protection, General data protection regulation (GDPR). https://gdpr-info.eu, 2018 (accessed 15 January 2024).
[42] M. Tzanou, Health data privacy under the GDPR, Taylor Francis Limited, London, 2020.