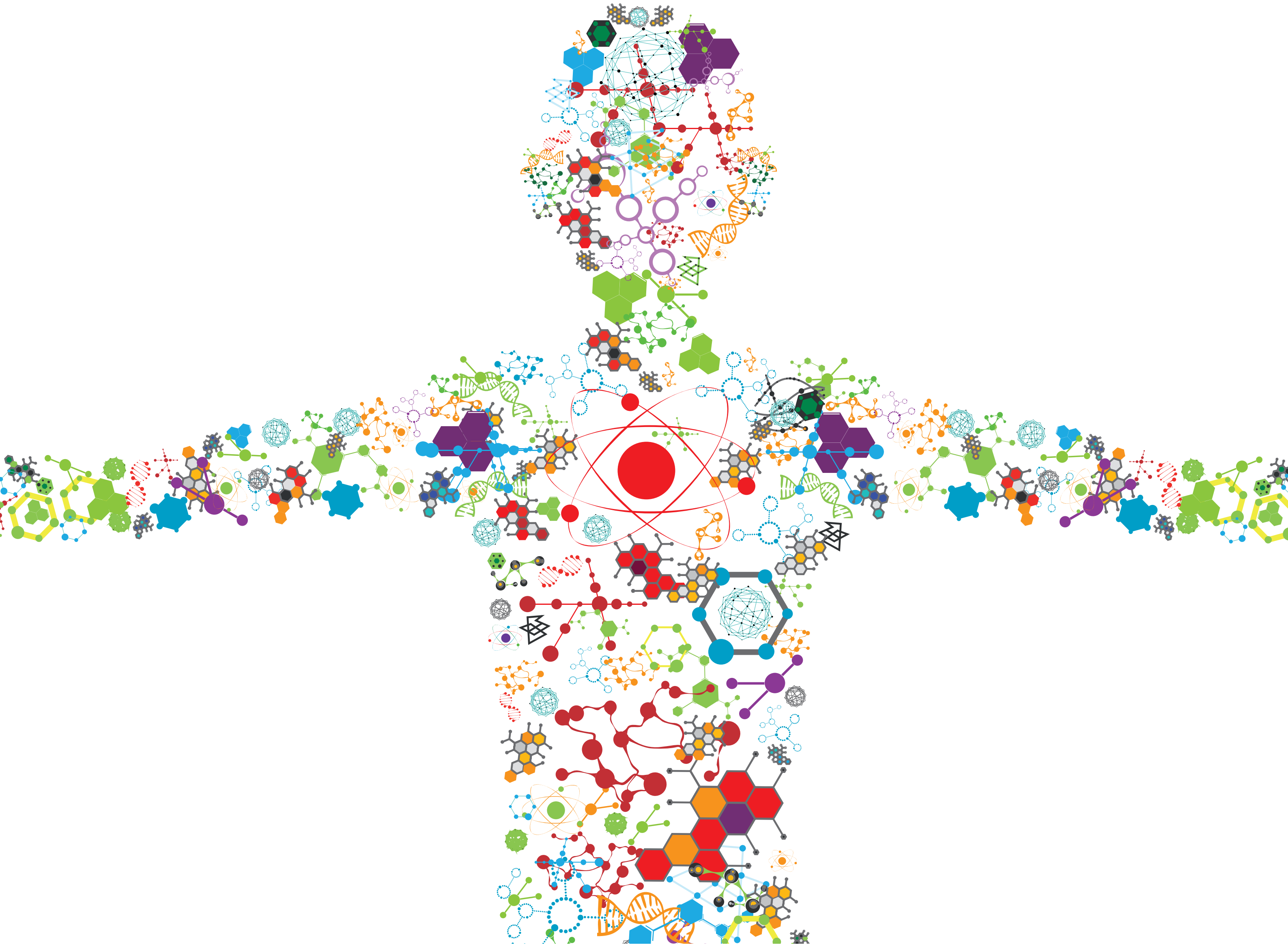
MAPPING THE CYBERBIOSECURITY ENTERPRISE

EDITED BY: Randall Murch and Diane DiEuliis

PUBLISHED IN: Frontiers in Bioengineering and Biotechnology



frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88963-213-8
DOI 10.3389/978-2-88963-213-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

MAPPING THE CYBERBIOSECURITY ENTERPRISE

Topic Editors:
Randall Murch, Virginia Tech, United States
Diane DiEuliis, National Defense University, United States

Citation: Murch, R., DiEuliis, D., eds. (2019). Mapping the Cyberbiosecurity Enterprise. Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-213-8

Table of Contents

frontiers
in Bioengineering and Biotechnology

EDITORIAL
published: 03 October 2019
doi: 10.3389/fbioe.2019.00235

Editorial: Mapping the Cyberbiosecurity Enterprise

Randall Murch[1*†] *and Diane DiEuliis*[2†]

[1] *Virginia Tech, Blacksburg, VA, United States,* [2] *National Defense University, Washington, DC, United States*

Keywords: cyberbiosecurity, bioeconomy, biosecurity, national security, biotechnology

Editorial on the Research Topic

Mapping the Cyberbiosecurity Enterprise

We are pleased to introduce this Research Topic in Frontiers in Bioengineering and Biotechnology on a new area of biosecurity, termed "Cyberbiosecurity." This term, originally introduced in the recently published strategic article by Murch et al. entitled "Cyberbiosecurity: An Emerging New Discipline to Help Safeguard the Bioeconomy (*Front. Bioeng. Biotechnol.* doi: 10.3389/fbioe.2018.00039), describes the security vulnerabilities that exist at the intersection of cybersecurity, cyber-physical security, and biosecurity.

Entitled "*Mapping the Cyberbiosecurity Enterprise,*" this collective of papers was amassed to firmly establish this topic as a new discipline within biosecurity. Each article contributes to developing and presenting deeper understanding of this emerging topic, and helps to delineate the range of current and potential applications of cyberbiosecurity. We also anticipate that this collective will foster greater engagement between the biosecurity and cybersecurity communities.

"Cyberbiosecurity" has been defined as "understanding the vulnerabilities to unwanted surveillance, intrusions, and malicious and harmful activities which can occur within or at the interfaces of comingled life and medical sciences, cyber, cyber-physical, supply chain and infrastructure systems, and developing and instituting measures to prevent, protect against, mitigate, investigate and attribute such threats as it pertains to security, competitiveness, and resilience." While cybersecurity is a broad and well-researched existing field, its application to specific aspects of the life sciences necessitates a conjoining of experts from each discipline which have predominantly existed in silos to date. Defining cyberbiosecurity as a discipline is a necessary first step in bringing these disparate groups together to expand understanding of the risks from their relative perspectives.

Mapping the topology of cyberbiosecurity has just begun, but proponents have realized that it has expansive applications across the life sciences, most obviously in the biomedical and pharmaceutical domains. But as the digitization of biology grows, biotechnology is expanding far beyond these traditional silos. The purposeful engineering of biology, including application of the classical "design, build, test" cycle, is opening unprecedented opportunities for biomaterials and biofuels and their use, for agriculture and food systems (from large scale crop engineering to "farm to table"), and for bioinformatics and "AI" (from small field tools to large-scale complex systems and cloud computing). As biotechnologies continue to advance and evolve, cyberbiosecurity will be a key consideration in existing critical infrastructure related to all these arenas. Further, new components of critical infrastructure may emerge and be defined through advances in the synthetic biology industry, and cybersecurity will need to be assessed for those new components. In our view, awareness and identification of vulnerabilities is an important first step in launching the field, followed by the development and implementation of mitigations and solutions. Eventually, practitioners in this growing field will be responsible for the development of guidelines and standards of governance, which will require adherence and compatibility with existing national defense strategies.

OPEN ACCESS

Edited and reviewed by:
Kenneth I. Berns,
University of Florida, United States

****Correspondence:***
Randall Murch
rmurch@vt.edu

[†]These authors have contributed equally to this work

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *30 August 2019*
Accepted: *10 September 2019*
Published: *03 October 2019*

Citation:
Murch R and DiEuliis D (2019) Editorial: Mapping the Cyberbiosecurity Enterprise. Front. Bioeng. Biotechnol. 7:235. doi: 10.3389/fbioe.2019.00235

This Special Collection, represented by both U.S. and international contributors, includes writings on a number of the topical areas described above. Vulnerabilities associated with synthetic biological manufacturing are described, including specific discussions of biopharmaceutical production. The evolving platforms for biotechnology, including distributed manufacturing models and laboratory automation, are included for consideration. Importantly, a discussion of the public health and stability ramifications of cyberbiosecurity in settings outside the US are also considered. General themes in other fields, such as agriculture, biopharma, and labs of the future are represented in stand-alone contributions. Some technical aspects of tool development, such as DNA synthesis security screens, and access to pathogen genome databases provide insights on current thinking and perceptions of risk. Finally, broad consideration is given to cyberbiosecurity in the national security context, given any new aspect of biosecurity must mesh with existing national security approaches and frameworks in the biodefense realm. Authors have also provided discussions of options for training and strategies for workforce development, all of which can help to build not only a general awareness of cybersecurity among biologists and synthetic biology engineers, but potentially develop a core of cyberbiosecurity specialists or practitioners that will be needed for risk assessments and solutions.

It is our hope that this eclectic set of insights and perspectives will broadly stimulate academia, government, non-profits, and the private sector to identify, prioritize, resource and pursue research, and implement solutions in the realm of cyberbiosecurity. Such research, outcomes and change management should focus on risk analysis, methods and technologies, education and training, guidelines and standards, policy, regulations and legal frameworks.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Bioengineering and Biotechnology

PERSPECTIVE
published: 06 June 2019
doi: 10.3389/fbioe.2019.00099

Cyberbiosecurity: A Call for Cooperation in a New Threat Landscape

Lauren C. Richardson [1]*, Nancy D. Connell [2], Stephen M. Lewis [1], Eleonore Pauwels [3] and Randy S. Murch [4]

[1] Merrick & Co., Arlington, VA, United States, [2] Johns Hopkins Center for Health Security, Bloomberg School of Public Health, Baltimore, MD, United States, [3] Wilson Center Science and Technology Innovation Program, The Wilson Center, Washington, DC, United States, [4] Virginia Tech Research Center, School of Public and International Affairs, Virginia Polytechnic Institute and State University, Arlington, VA, United States

OPEN ACCESS

Edited by:
Stephen Allen Morse,
Centers for Disease Control and Prevention (CDC), United States

Reviewed by:
Dana Perkins,
United States Department of Health and Human Services, United States
Laura Adam,
Ebiosec, Inc, United States

***Correspondence:**
Lauren C. Richardson
Lauren.Richardson@merrick.com

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 30 January 2019
Accepted: 18 April 2019
Published: 06 June 2019

Citation:
Richardson LC, Connell ND, Lewis SM, Pauwels E and Murch RS (2019) Cyberbiosecurity: A Call for Cooperation in a New Threat Landscape.
Front. Bioeng. Biotechnol. 7:99.
doi: 10.3389/fbioe.2019.00099

The life sciences now interface broadly with information technology (IT) and cybersecurity. This convergence is a key driver in the explosion of biotechnology research and its industrial applications in health care, agriculture, manufacturing, automation, artificial intelligence, and synthetic biology. As the information and handling mechanisms for biological materials have become increasingly digitized, many market sectors are now vulnerable to threats at the digital interface. This growing landscape will be addressed by cyberbiosecurity, the emerging field at the convergence of both the life sciences and IT disciplines. This manuscript summarizes the current cyberbiosecurity landscape, identifies existing vulnerabilities, and calls for formalized collaboration across a swath of disciplines to develop frameworks for early response systems to anticipate, identify, and mitigate threats in this emerging domain.

Keywords: biosecurity, cybersecurity, cyberbiosecurity, life sciences, bioeconomy, bioinformatics, synthetic biology, biomanufacturing

INTRODUCTION

The greatest vulnerabilities in any field can be found at its margins—at its junctions with adjacent fields. The new discipline of cyberbiosecurity has been created to bring together disparate communities to identify and address a complex ecosystem of security vulnerabilities at the interface of the life sciences, information systems, biosecurity, and cybersecurity (Murch et al., 2018; Peccoud et al., 2018); it serves as a lens for observation that relies on disciplinary integration. Cyberbiosecurity describes an intersection of disciplines that falls outside any single sector; because these convergences are not clearly analyzed, actors within a single sector do not have agency to address potential issues and are less likely to cooperate. Such vulnerabilities exist within biomanufacturing, cyber-enabled laboratory instrumentation and patient-focused systems, "Big Data" generated from "omics" studies, and throughout the farm-to-table enterprise (**Figure 1**). In addition to fundamental and applied research and development opportunities, off-the-shelf solutions not yet applied in this domain likely exist. While the term is new, the concept of

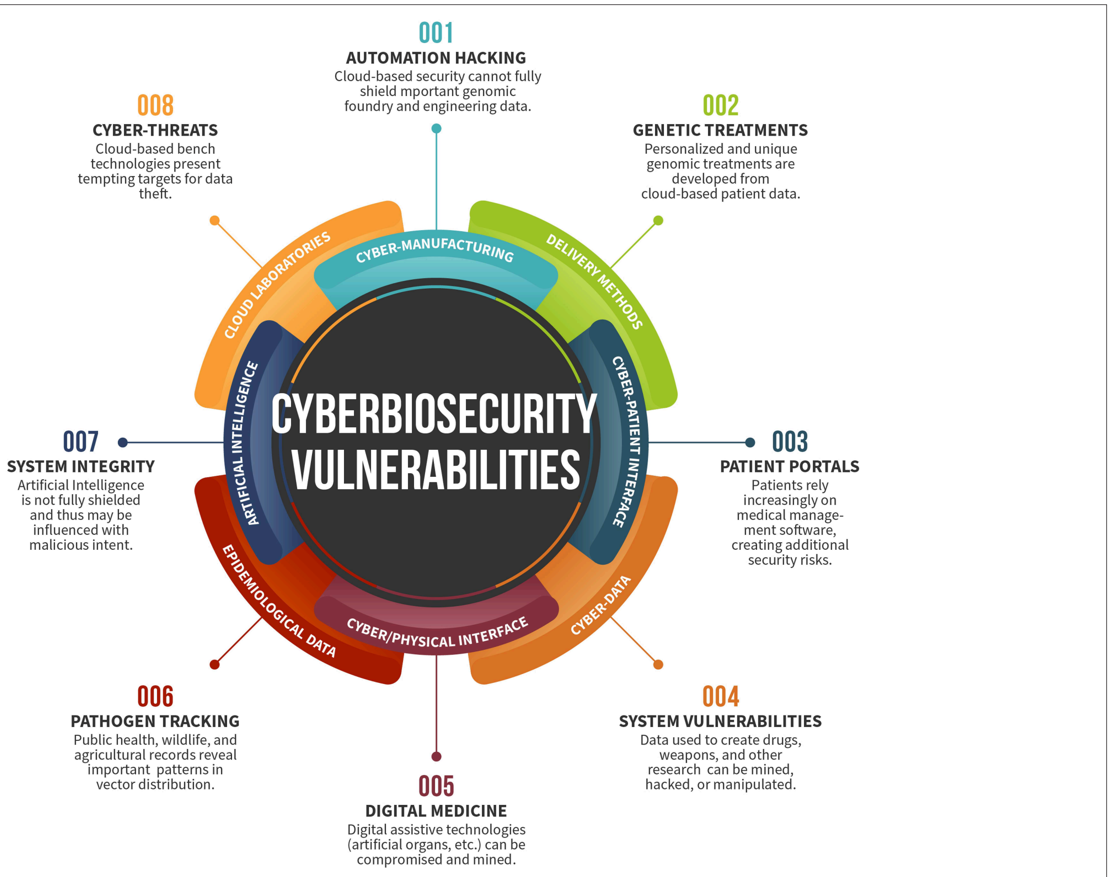

001
AUTOMATION HACKING
Cloud-based security cannot fully shield mportant genomic foundry and engineering data.
002
GENETIC TREATMENTS
Personalized and unique genomic treatments are developed from cloud-based patient data.
003
PATIENT PORTALS
Patients rely increasingly on medical management software, creating additional security risks.
004
SYSTEM VULNERABILITIES
Data used to create drugs, weapons, and other research can be mined, hacked, or manipulated.
005
DIGITAL MEDICINE
Digital assistive technologies (artificial organs, etc.) can be compromised and mined.
006
PATHOGEN TRACKING
Public health, wildlife, and agricultural records reveal important patterns in vector distribution.
007
SYSTEM INTEGRITY
Artificial Intelligence is not fully shielded and thus may be influenced with malicious intent.
008
CYBER-THREATS
Cloud-based bench technologies present tempting targets for data theft.
CYBER-MANUFACTURING
DELIVERY METHODS
CYBER-PATIENT INTERFACE
CYBER-DATA
CYBER/PHYSICAL INTERFACE
EPIDEMIOLOGICAL DATA
ARTIFICIAL INTELLIGENCE
CLOUD LABORATORIES
CYBERBIOSECURITY
VULNERABILITIES

FIGURE 1 | Summary of cyberbiosecurity vulnerabilities across the spectrum of the bioeconomy. Each sector of the wheel describes specific vulnerabilities that span the medicine, infectious disease, systems management, and biotechnology.

cyberbiosecurity has been acknowledged as a serious concern (Wintle et al., 2017). The issues raised in the area of cyberbiosecurity will have substantial impact on the growing bioeconomy[1].

The solution set is not simply technical: creating cross-sector convergence opportunities for effective communication and collaboration as well as governance, policy, and regulatory structures is also necessary. Derived value from cyberbiosecurity endeavors potentially embraces economic impact, national security, societal resilience, and environmental sustainment. In this paper, we establish a landscape for cyberbiosecurity and issue a call for cooperation across sectors to recognize and mitigate potential threats.

BACKGROUND

As a part of the discussion, we refine the definition of cyberbiosecurity. **Cybersecurity** encompasses the protection of computer systems from theft and damage to their hardware, software, or information, as well as from disruption or misdirection of the services they provide. **Biosecurity** involves securing valuable biological material from misuse or harm. Initially, Murch et al. defined cyberbiosecurity as the "developing understanding of the vulnerabilities to unwanted surveillance, intrusions, and malicious and harmful activities which can occur within or at the interfaces of comingled life science, cyber, cyber-physical, supply chain and infrastructure systems, and developing and instituting measures to prevent, protect against, mitigate, investigate, and attribute such threats as it pertains to security, competitiveness, and resilience" (Murch et al., 2018). The definitions of cybersecurity and biosecurity both include an underlying assumption of value on the part of the material in question. We further suggest expansion of this definition of cyberbiosecurity to differentiate it from the individual scopes of cybersecurity and biosecurity. Cyberbiosecurity addresses the potential for or actual malicious destruction, misuse, or exploitation of valuable information, processes, and material at the interface of the life sciences and digital worlds; concept mastery requires an understanding of this interface in the context of the threat of malignant use of technology in general. This paper is a call to action before such a succession of events takes place.

[1] Bioeconomy is defined as "economic activity that is fueled by research and innovation in the biological sciences (House, 2012)."

LANDSCAPE

Cyberbiosecurity cuts across disciplines; impacting fields from laboratory science, to human and animal health, agriculture, and environmental health and ranging from protection to management and remediation. Technology integration is the new norm, with novel technology improvements and simple digitization bringing easy access to old systems, such as medical records. As technical disciplines develop at an exponential pace and their convergence accelerates, it is becoming increasingly clear that the fields of cybersecurity and biosecurity must also converge in order to address inherent digital and biological concerns. Further, technological convergence meets the decreasing cost for access at the Do It Yourself (DIY)/community biology space.

CYBERBIOSECURITY IN BIOTECHNOLOGY

Artificial Intelligence

Industry interest in artificial intelligence (AI) has experienced a resurgence in recent years due to increased computing power, advancing applications of neural networks, and an emergence of new machine and deep learning techniques across the biology sector. Biotechnology companies are successfully utilizing these developments for drug design and development (Zilinskas, 2017), genomics (Pauwels and Vidyarthi, 2017), evolutionary biology (Feltes et al., 2018), protein folding (Paladino et al., 2017), and more. This rapid and evolving interest in the landscape of new AI technologies has led to emerging threat domains related to information privacy and storage, ownership over biological and genetic data, and applications of powerful technologies (Pauwels, 2018). These issues are not new, as bioinformatics and digitization have created a potential target; however, the popularization of AI has refreshed these concerns in the modern zeitgeist. There is a renewed opportunity for life science and cybersecurity professionals to design and implement frameworks to facilitate responsible application of AI techniques to biology.

Automation

The convergence of robotics, machine learning, and artificial intelligence has paved the way for automated approaches to biology, manufacturing, software development, accounting, and more. Improved biological engineering techniques and robotics have converged to result in rapid prototyping and higher yields. Laboratories are increasingly using robots to improve throughput and free up the hands of laboratorians around the world (McGee, 2014; Szesterniak, 2014). As robots are increasingly connected to networks and other electronic systems, new cyberbiosecurity concerns unique to automated laboratory environments are beginning to emerge. Virtual environments allow access to infrastructure within the physical world; this creates a vulnerability that would permit unauthorized remote access to an automated biological manufacturing system. As automation increases within the life sciences, so too will potential vulnerabilities to threat.

Synthetic Biology

The term “synthetic biology” is widely used to describe activities carried out by scientists in a variety of disciplines, from bioengineering, chemistry, biochemistry, and materials science to cellular and molecular biology (Hobom, 1980; Purnick and Weiss, 2009). Today, engineers, biologists, technologists, and citizen scientists have turned this field into a true discipline. Systems engineering techniques are being applied to organisms to design genetic circuits, novel molecules, and commodities such as fuels, electricity, feed, and renewable materials (Rollin et al., 2013; Kiss et al., 2014). Simultaneously, the design-build-test approach traditionally used in product development is rapidly emerging in organism engineering (Dudley et al., 2015; Gill et al., 2016). Advancements in synthetic biology will have a significant impact on cyberbiosecurity as laboratory automation techniques become more widespread and the traditional cost barrier for scale-up of production is lowered. Similarly, the convergence of robotics, microfluidics, cell-free systems design and synthetic metabolic engineering stands to create new cyberbiosecurity risks and unique threat domains (Nielsen and Keasling, 2011; Murch et al., 2018; Peccoud et al., 2018). As these fields further develop and converge, revealed vulnerabilities will offer new opportunity for exploitation.

CYBERBIOSECURITY IN DIGITIZATION OF TRADITIONAL TECHNOLOGY

Manufacturing

Science and technology-reliant organizations are becoming more complex and networked throughout facilities, supply chains, logistics, and transport mechanisms. Distributed manufacturing employs decentralized production networks linked by information technology; as more connections between traditionally isolated systems are developed, more security controls must be considered in order to mitigate risks and reduce vulnerabilities. The production processes and assemblies of biologics and other materials can also be distributed and carried out asynchronously at geographically different locations, allowing response to potential threats to be developed *in situ*.

In addition to facilitation of distributed manufacturing techniques for traditional life sciences operations, recent advances in cell-free metabolic engineering technologies allow for higher throughput in production environments. This has resulted in improved biological techniques for rapid prototyping and higher yields. Cell-free biological systems are being used to develop commodities such as fuels, electricity, feed, and renewable materials (Rollin et al., 2013). As the convergence of dichotomous technical disciplines (e.g., automation and cellular biology) continues to expand rapidly, it is increasingly important that the fields of cybersecurity and biosecurity converge to address inherent digital and biological concerns.

Biomedical Sciences

Cybersecurity and health security converge with increasing digitization of health data. Regulatory mechanisms are in place to address concerns regarding privacy and confidentiality of medical and billing information; however, this extends beyond the cyber-patient interface in the context of electronic medical records. Patient treatment management—including potential drug interactions, protocols, and sensitivities specific to the patient—is increasingly digitized. Personalized medicine diagnostics and therapeutics are rapidly expanding, and much of the information associated with these interventions is maintained digitally. Biomedical data breaches are not without historic precedent: in 2014, data breaches of three major health systems resulted in unauthorized access to millions of patient records, including clinical data (Kozminski, 2015). These breaches provided the perpetrators valuable clinical data, which could be used internally or sold for monetary gain. In addition to facilitating illicit data collection, disruption of digitally-programmed diagnostic testing systems or therapeutic targeting fields could result in ineffective treatment. Medical devices are also an area of interest in cyberbiosecurity, as many potential exploits could be leveraged through direct and indirect interfaces with the patient and manufacturer (Khera, 2017).

Agriculture

Throughout much of the world, food and beverage safety and security is a high priority. Concomitantly, the economics, societal robustness, and security implications of agriculture, foodstuffs and beverages are massive. Extensive quality measures are in place to prevent and mitigate threats from manifesting; outbreak and contamination detection and response systems react when problems are noticed. Packaging and labeling methodology have also been improved. However, agriculture and consumables in many countries rely on cyber-enabled systems for many aspects of farm management, production-to-consumption, raw materials to finished product, and logistics (Security Security DoH., 2018). The health and security of this dimension of agriculture and food systems is unclear from a cyberbiosecurity perspective. We reason that vulnerable critical links and nodes exist throughout this highly complex global and national ecosystem; attention to cyberbiosecurity measures is warranted and would be considerably beneficial.

CONCLUSION

The convergence of recent advances in the life sciences with regard to traditional cybersecurity threats has led to the recognition and identification of vulnerabilities, known as cyberbiosecurity threats (Murch et al., 2018; Peccoud et al., 2018). Here we present a preliminary review of the landscape of these threats and propose recommendations to activate a "call to action" to anticipate these threats and mitigate their effects. Several entities have approached related issues: for example, in October 2019, HHS announced the opening of the Health Sector Cybersecurity Coordination Center (HC3), intended to prevent threats to health data through strengthening cybersecurity (Office Office HP., 2018). Though concurrent efforts touch on the issues described, individual efforts alone are insufficient to cover the breadth of the landscape. We call for analyses and publications to fully scope cyberbiosecurity and identify a comprehensive strategy to establish the discipline's goals and objectives; we call for carefully-crafted national or international meetings of experts from appropriate science, technology, and social science domains to begin to bring communities together to define priorities for approaches to solutions by examining causes, effects and possible remedies; we call for initiation of campaigns of blended teams of experts engaging key government agencies to raise awareness and initiate creation of and/or changes to relevant policies and programs in order to incorporate relevant cyberbiosecurity perspectives.

AUTHOR CONTRIBUTIONS

LR, NC, SL, EP, and RM contributed conception and design of the manuscript. LR, NC, SL, and RM wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

The authors acknowledge the graphic talents of LJ DeGraffenreid.

REFERENCES

Dudley, Q. M., Karim, A. S., and Jewett, M. C. (2015). Cell-free metabolic engineering: biomanufacturing beyond the cell. *Biotechnol. J.* 10, 69–82. doi: 10.1002/biot.201400330

Feltes, B. C., Grisci, B. I., Poloni, J. F., and Dorn, M. (2018). Perspectives and applications of machine learning for evolutionary developmental biology. *Mol. Omics* 14, 289–306. doi: 10.1039/C8MO00111A

Gill, R. T., Halweg-Edwards, A. L., Clauset, A., Way, S. F., et al. (2016). Synthesis aided design: the biological design-build-test engineering paradigm? *Biotechnol. Bioeng.* 113, 7–10. doi: 10.1002/bit.25857

Hobom, B. (1980). Gene surgery: on the threshold of synthetic biology. Med. *Klin.* 75, 834–841.

House, T. W. (2012). National bioeconomy blueprint, April 2012. *Industrial Biotechnol.* 8, 97–102. doi: 10.1089/ind.2012.1524

Khera, M. (2017). Think like a hacker: insights on the latest attack vectors (and security controls) for medical device applications. *J. Diabetes Sci. Technol.* 11, 207–212. doi: 10.1177/1932296816677576

Kiss, A. A., Grievink, J., and Rito-Palomares, M. (2014). A systems engineering perspective on process integration in industrial biotechnology. *J. Chem. Tech. Biotech.* 90, 349–355. doi: 10.1002/jctb.4584

Kozminski, K. G. (2015). Biosecurity in the age of Big Data: a conversation with the FBI. *Mol. Biol. Cell* 26, 3894–3897. doi: 10.1091/mbc.E14-01-0027

McGee, J. (2014). Screening Robotics and Automation. *SLAS Discov. Adv. Life Sci.* 19, 1131–1132. doi: 10.1177/1087057114538231

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

Nielsen, J. K., and Keasling, J. D. (2011). Synergies between synthetic biology and metabolic engineering. *Nat. Biotechnol.* 29, 693–695. doi: 10.1038/nbt.1937

Office HP. (2018). *HHS Announces the Official Opening of the Health Sector Cybersecurity Coordination Center*. U.S. Department of Health & Human Services: HHS.gov.

Paladino, A., Marchetti, F., Rinaldi, S., and Colombo, G. (2017). Protein design: from computer models to artificial intelligence. *WIREs Comput. Mol. Sci.* 7:e1318. doi: 10.1002/wcms.1318

Pauwels, E. (2018). *The Ethical Anatomy of Artifical Intelligence*. New York, NY: U.N. University.

Pauwels, E., and Vidyarthi, A. (2017). *Who Will Own the Secrets in Our Genes? A US-China Race in Artificial Intelligence and Genomics*. Washington, DC: Woodrow Wilson International Center for Scholars.

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2018). Cyberbiosecurity: from naive trust to risk awareness. *Trends Biotechnol.* 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012

Purnick, P. E., and Weiss, R. (2009). The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* 10, 410–422. doi: 10.1038/nrm2698

Rollin, J. A., Tam, T. K., and Zhang, Y. H. P. (2013). New biotechnology paradigm: cell-free biosystems for biomanufacturing. *Green Chem.* 15, 1708–1719. doi: 10.1039/c3gc40625c

Security DoH. (2018). *Threats to Precison Agriculture*.

Szesterniak, M. (2014). Six *Trends in Robotics in the Life Sciences*. Available online at: http://www.parkermotion.com/whitepages/Six-Trends-in-Life-Science-Robotics.pdf

Wintle, B. C., Boehm, C. R., Rhodes, C., Molloy, J. C., Millett, P., Adam, L., et al. (2017). Point of view: a transatlantic perspective on 20 emerging issues in biological engineering. *Elife* 2017:e30247. doi: 10.7554/eLife.30247

Zilinskas, R. A. (2017). A brief history of biological weapons programmes and the use of animal pathogens as biological warfare agents. *Rev. Sci. Tech.* 36, 415–422. doi: 10.20506/rst.36.2.2662

Conflict of Interest Statement: LR and SL were employed by Merrick and Company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

BRIEF RESEARCH REPORT
published: 19 June 2019
doi: 10.3389/fbioe.2019.00136

Cyber-Biosecurity Risk Perceptions in the Biotech Sector

Kathryn Millett, Eduardo dos Santos and Piers D. Millett*

Biosecure Ltd, Market Rasen, United Kingdom

The expanding digitization of the biological sciences places greater value on the data generated, information extrapolated and knowledge gained. Failing to protect data will affect a company or country's ability to position itself optimally in the forthcoming fourth industrial revolution. Further, more reliance on automation, distribution, and outsourcing in biotechnology makes its infrastructure a target. The equipment and service providers that drive physical research and development are also all connected online. Failing to protect these resources from intrusion increases the risk of accidental or deliberate harm, for example by the loss of control over biological products. Robust cybersecurity measures are therefore critical for both securing the data generated by the biotechnology sector as well as securing key infrastructure. Cyber-biosecurity is emerging multidisciplinary field that combines cybersecurity, biosecurity, and cyber-physical security as relates to biological systems (Murch et al., 2018). To better identify the perceived risks at the interface between cybersecurity and biosecurity, Biosecure conducted a pilot study that surveyed the opinions of a discrete set of international field leaders in biotechnology and cybersecurity. The survey was carried out online from October-November 2017. Key findings of the survey showed that cyber-biosecurity risks were considered to be difficult to characterize due to variations in types of threats, targets and potential impacts, and compounded by a notable variation between the level of sophistication or maturity of mitigation and response measures. Further research is therefore necessary bringing together the different communities focusing on these issues to develop a common language, better define the threats and discuss potential ways forward in addressing risks.

Keywords: cyber-biosecurity, biotechnology, bioeconomy, infrastructure, risk perception, biosecurity, industry

OPEN ACCESS

Edited by:
Diane DiEuliis,
National Defense University,
United States

Reviewed by:
Gregory D. Koblentz,
George Mason University,
United States
Dana Perkins,
United States Department of Health and Human Services, United States

****Correspondence:***
Kathryn Millett
kathryn@biosecu.re

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *26 November 2018*
Accepted: *21 May 2019*
Published: *19 June 2019*

Citation:
Millett K, dos Santos E and Millett PD (2019) Cyber-Biosecurity Risk Perceptions in the Biotech Sector. Front. Bioeng. Biotechnol. 7:136. doi: 10.3389/fbioe.2019.00136

INTRODUCTION

The development and recognition of "cyber-biosecurity" as an important element in securing data and products emerging from the biotechnology and biomedical sectors has predominantly emerged from the field of biosecurity. While the risks relating to accessing private biomedical data and the theft of valuable data from an intellectual property standpoint are well-known and recognized, the biosecurity implications of cyber intrusions relating to biotechnology infrastructure remain largely unknown in commercial biotechnology facilities.

To better gauge the current level of understanding and awareness of cyber-biosecurity risks in the biotechnology sector and identify how the risks are perceived, Biosecure conducted a pilot survey targeting a discrete set of international leaders in the fields of biotechnology and cybersecurity.

METHODOLOGY

To conduct a discrete pilot survey of the types and level of cyber-biosecurity risks identified in the field of biotechnology, a short questionnaire comprising 12 questions that was posted securely online. The questions posed were a mix of multiple choice and open-ended questions, divided across the themes of risk perception and awareness, risk mitigation capacities and resources, and the urgency of, and potential avenues for, any future action. The questions were reviewed by an expert in qualitative methodology to eliminate any issues of bias.

The survey described in this paper was conducted in accordance with the Declaration of Helsinki and all participants provided informed consent in writing (World Medical Association, 2013). The survey described is not considered research by the UK National Health Service and Medical Research Council and does not require review by a Research Ethics Committee. In addition, Biosecure Ltd. funded the survey using its own corporate funds. Biosecure Ltd. does not, and has not, received US Federal research funding. As a result, the survey described in this paper was performed in accordance with relevant institutional and national guidelines.

Twenty-six individuals were invited to participate from across the biotech and cybersecurity sector. Invitees from the biotech sector included founders of small to medium biotechnology companies in the United States and United Kingdom, senior management of large biotechnology companies (with an international footprint), representatives of industry, venture capitalists specializing in biotechnology, and advisors to the above on security issues. The individuals approached in the cybersecurity sector included industry specialists, leading academics, national government experts, experts in leading think tanks, and specialists within intergovernmental organizations.

Overall, of the 26 invited questionnaire participants, 13 agreed to participate. The responses were anonymized.

SURVEY RESULTS

The results of the survey were assessed according to four key areas: (1) assessing the threat; (2) assessing threat mitigation and response capacity; (3) available tools and resources; and, (4) recommended next steps. The key findings under each of these areas are elaborated below and summarized in **Table 1**.

Assessing the Threat

Over two-thirds of respondents deemed the risks posed to the biotechnology sector by cyber threats and intrusions as elevated or severe when compared to normal operating standards in the biotech industry. The two scenarios perceived to pose the greatest risk were: unauthorized access to data, information, or knowledge outside the public domain; and unauthorized actors able to secretly change data, information, or knowledge. In only one scenario (in which an unauthorized actor takes control of infrastructure) did any respondent think there was no or minimal risk.

- When asked to identify different types of risks from cybersecurity breaches in the biotech sector, participants noted potential negative impacts from:
- The theft, elimination or ransom of data, algorithms, or software with a direct or indirect impact on R&D or commercial operations;
- Modification of data, algorithms, or software with a direct or indirect impact on research and development or commercial operations;
- The loss of intellectual property or commercial advantage by data, algorithms, or software being available to competitors;
- Potential for the disabling or disruption of important systems or infrastructure leading to disruption of commercial operations or impeding good manufacturing practices;
- Manipulation of bio-manufacturing or automated systems to create risks.

Respondents ranked states and proxies used by states as the type of actor posing the greatest risk, with lone individuals viewed as generating the least risk. This survey did not differentiate between insider or outsider threats, regardless of whether states, groups or lone individuals. This may be an area ripe for further study.

All participants considered that cyber-biosecurity risks posed a real and current threat, but that these were not, or only partially, being addressed within the biotech sector. In part, this was considered due to a lack of awareness and information within the biotech community, with one participant noting that "[M]any companies are unaware of the intensity of outsider threats because they are not actively monitoring these activities."

Assessing Current Threat Mitigation and Response Capacities

While noting the lack of sufficient information on the type and level of biorisks to the biotech sector by cyber intrusions, over seventy-five per cent (75%) of participants indicated that their organizations had undertaken some efforts to address cybersecurity issues, and ninety per cent (90%) of these reported that such measures were regularly reviewed.

However, the comprehensiveness and maturity of mitigation efforts were reported as being varied, with some participants reporting that their efforts were only in the nascent stages. One respondent, for example, noted that their activities had been "...mostly discussions that it will be a problem but they have no idea nor urge to address it." Another noted that the issues had been considered "[F]airly deeply, although [we] have not… done any work to implement anything."

By contrast, other participants had begun integrating cybersecurity into their business with a participant reporting that "[W]e have considered security implications in our technology development at all levels… partner technologies we integrate have always required a careful discussion of the security implications that flow from their use, and as a result we rely heavily on technologies from vendors such as Google and Microsoft that have strong security cultures."

In addition to variances in awareness and the perceived risks posed by cyber-attacks to biological facilities and equipment, respondents pinpointed the lack of available resources as a limiting factor for addressing cyber-biosecurity. Over ninety per cent (90%) of participants expressed a strong view that insufficient time and resources are being dedicated to dealing with these risks. One participant noted they "have not yet had the resources to do formal red team testing of our systems" and another commented that "[S]ufficient time and resources are almost never dedicated to dealing with risks from cybersecurity; biotech is no exception." Further, it was remarked that "[D]ealing with cybersecurity breaches is not a one size fits all process. Filling the gaps on the topic requires a tailored approach for each company, entity, or facility. By performing a comprehensive gap analysis for each entity, the answer to this question can be discovered."

When asked their view on the appropriate agency to take the lead in addressing any risks from cybersecurity breaches in biotechnology, participants showed a wide divergence of opinion (**Figure 1**) suggesting that a multi-stakeholder approach may be warranted.

Available Tools and Resources

Over seventy-five (75%) of respondents were unaware of any dedicated resources (reports, guidance, standards, etc.) for dealing with risks from cybersecurity breaches in biotechnology. Those that were aware of existing resources highlighted internal company resources, broader standards that incorporated aspects of biosecurity and cybersecurity but which did not specifically address the overlap, or country-specific resources, such as National Institute of Standards and Technology and FBI outreach agents in the USA.

However, there was greater awareness (50%) of the existence of "dedicated support for dealing with this issue (such as hotlines, reporting infrastructure, national experts, commercial services,

TABLE 1 | Relative risk perception of different cybersecurity threats to biotech.

	No or minimal risk	Risk comparable to normal operating standards	Elevated or severe risk
An incident in which an unauthorized actor takes control of infrastructure (e.g., lab equipment, lab control systems, or even a fully automated robot lab)	2	2	9
An incident in which an unauthorized actor accesses data, information, or knowledge that is not in the public domain	0	2	11
An incident in which an unauthorized actor is able to circumvent security controls, such as those used to screen orders and customers amongst certain biotech service providers	0	3	9
An incident in which an unauthorized actor is able to secretly change data, information, or knowledge	0	1	12
An incident in which an unauthorized actor is able to interrupt the functioning of lab systems	0	4	9
An incident originating from a compromise in the supply chain	0	2	9

White, No response; Yellow, 1 to 5 responses; Orange, 6 to 10 responses; Red, Over 10 responses.

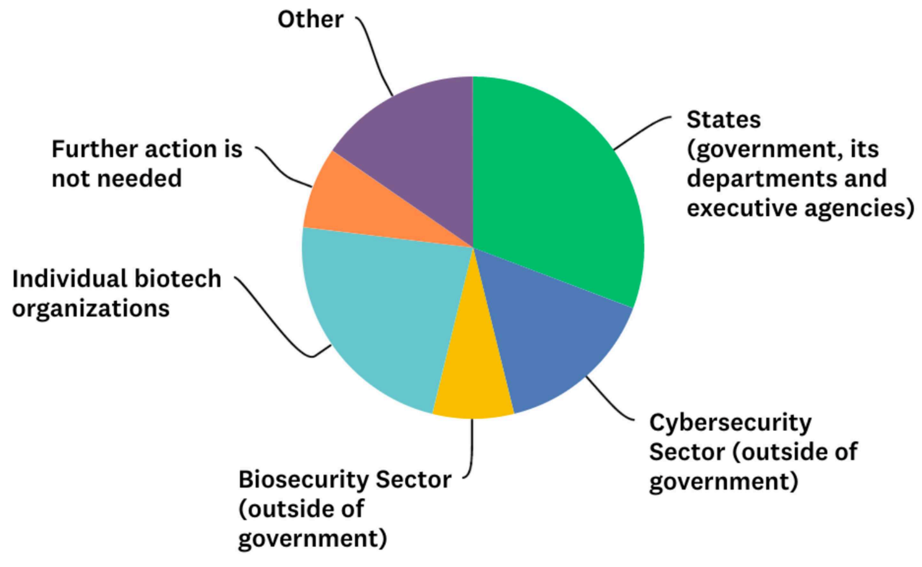


FIGURE 1 | Views as to the appropriate primary actor in addressing any risks from cybersecurity breaches in biotech.

etc." with two thirds of those respondents aware of support citing the Weapons of Mass Destruction (WMD) Directorate of the FBI and one respondent citing private company, Ebiosec. No participant identified sources of support that specifically address the cybersecurity needs of the biotech sector outside of the USA.

Recommended Next Steps in Addressing Cyber-Biosecurity

Several respondents pointed to efforts to address gaps in the interface between cyber- and biosecurity including sponsored meetings and, in a few cases, having specifically allocated staff time to addressing these issues. In addition, notice has been made of the emergence of new actors in the field, including such as companies like Ebiosec which provides services to "manage, model, secure, and visualize their data-driven life sciences operations[1]." The founders of this company also manage an online portal for "fostering discussions and sharing information, events and tools to secure the digital dimension of the biothreat[2]."

However, the majority of participants acknowledged that much more needs to be done to bring together the communities addressing biosecurity and cybersecurity, and identify effective measures and approaches to mitigate and prevent the risks, including fine tuning broader regulatory approaches to help foster a cybersecurity culture. One participant noted "Biotech does not think about security other than more traditional biosecurity and biosafety; security communities do not understand biotech (focused on traditional telecoms and digital)."

A number of issues warranting increased attention were also identified, including: the implications of new supply/value chains; techno-espionage or potential for business model/regulatory disruptions; loss of public/political trust resulting from inactivity; and how cybersecurity risk impacts competitiveness of biotechnology companies.

CONCLUSION

The issue of cyber-biosecurity is not well-known or understood, even among biotechnology and cybersecurity experts. A concerted effort to develop this emerging field, define, and foster awareness of the threats and craft a common language is therefore a pressing need as the digital age of biology progresses.

Opportunities are needed to bring together communities focusing on these issues, and begin work on areas of common interest and the means to address the identified risks. Strengthened multi-stakeholder capacity is needed to work at the interface between cybersecurity and biosecurity, and support and resources should be invested in further understanding cybersecurity risks in the biotechnology sector in order to develop appropriate counter measures.

AUTHOR CONTRIBUTIONS

KM is the lead author of this paper. PM and KM devised and carried out the survey. EdS provided technical assistance during the survey and conducted a literature review on cyberbiosecurity.

[1] See http://ebiosec.com/

[2] See http://information-biosecurity.org/

REFERENCES

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an Emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

World Medical Association (2013). *Declaration of Helsinki Ethical Principles for Medical Research Regarding Human Subjects*. Available online at: https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/

Conflict of Interest Statement: KM and PM are founders and owners of Biosecure Ltd. Biosecure Ltd. does not, and has not, received US Federal research funding.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer GK declared a past collaboration with one of the authors, PM, to the handling editor.

PERSPECTIVE
published: 25 February 2019
doi: 10.3389/fbioe.2019.00021

National and Transnational Security Implications of Asymmetric Access to and Use of Biological Data

Kavita M. Berger[1]* and *Phyllis A. Schneck*[2]

[1] Gryphon Scientific, LLC, Takoma Park, MD, United States, [2] Promontory Financial Group, an IBM Company, Washington, DC, United States

Biology and biotechnology have changed dramatically during the past 20 years, in part because of increases in computational capabilities and use of engineering principles to study biology. The advances in supercomputing, data storage capacity, and cloud platforms enable scientists throughout the world to generate, analyze, share, and store vast amounts of data, some of which are biological and much of which may be used to understand the human condition, agricultural systems, evolution, and environmental ecosystems. These advances and applications have enabled: (1) the emergence of data science, which involves the development of new algorithms to analyze and visualize data; and (2) the use of engineering approaches to manipulate or create new biological organisms that have specific functions, such as production of industrial chemical precursors and development of environmental bio-based sensors. Several biological sciences fields harness the capabilities of computer, data, and engineering sciences, including synthetic biology, precision medicine, precision agriculture, and systems biology. These advances and applications are not limited to one country. This capability has economic and physical consequences, but is vulnerable to unauthorized intervention. Healthcare and genomic information of patients, information about pharmaceutical and biotechnology products in development, and results of scientific research have been stolen by state and non-state actors through infiltration of databases and computer systems containing this information. Countries have developed their own policies for governing data generation, access, and sharing with foreign entities, resulting in asymmetry of data sharing. This paper describes security implications of asymmetric access to and use of biological data.

Keywords: biotechnology, cybersecurity, information security, data vulnerability, biological data, biosecurity, data access, data protection

OPEN ACCESS

Edited by:
Randall Steven Murch,
Virginia Tech, United States

Reviewed by:
Segaran P. Pillai,
United States Department of Homeland Security, United States
Jacqueline Fletcher,
Oklahoma State University, United States

****Correspondence:***
Kavita M. Berger
kberger@gryphonscientific.com

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *29 November 2018*
Accepted: *29 January 2019*
Published: *25 February 2019*

Citation:
Berger KM and Schneck PA (2019) National and Transnational Security Implications of Asymmetric Access to and Use of Biological Data. Front. Bioeng. Biotechnol. 7:21. doi: 10.3389/fbioe.2019.00021

INTRODUCTION

Advances in computer science, engineering, and data science have changed research, development, and application of biology and biotechnology in the United States and internationally. Examples of changes include: (a) increased reliance on internet connectivity for research and laboratory operations (Accenture, 2015; Bajema et al., 2018; Olena, 2018); (b) increased use of automation in life-science laboratories (Chapman, 2003); (c) application of the "design-build-test" paradigm to create new biological organisms (Agapakis, 2014; Carbonell et al., 2018); (d) increased generation, analyses, and computational modeling of information about biological systems, cells, and molecules (Thurow et al., 2004; Walpole et al., 2013); (e) treatment of organisms and DNA as materials rather than phenomena to study (Service, 2017; Anderson et al., 2018; Patel, 2018); and (f) new funders such as venture capital, crowdfunding platforms, and foreign companies and governments (Von Krogh et al., 2012; Cha, 2015; Mervis, 2017). These changes have transformed the scientific, agricultural, and health communities' ability to understand and manipulate the world around them. In addition, the changes have enabled an influx of new practitioners and problem-solvers into biology, providing opportunities for education and research all over the world.

Biotechnology harnesses the capabilities of computer, data, and engineering sciences to establish and advance new fields such as synthetic biology, precision medicine, precision agriculture, and systems biology. Cloud-based platforms and open source, easy-to-use software enable scientists from anywhere in the world to use advanced data analytics in their studies. The software and hardware emerging from these fields improve our collective understanding of molecular and systems-level genetics, new drug therapies for longer and better quality of life, and design of novel and/or unnatural organisms. Critical to these pursuits is the sharing of research results and underlying data, without which societal decision-making about human, animal, plant, and environmental health cannot be realized fully. However, during the past two decades, concerns about data sharing have been raised, resulting in the issuance of international, regional, and national-level policies governing access to different types of data, including biological data. In addition, the platforms through which data are stored, transported, and analyzed may be vulnerable to unauthorized acquisition of information by malicious actors, which could lead to significant economic and physical harms to the health, safety, and security of a population. Although not considered "dual use life sciences research of concern" U. S. Government, 2012, 2014), the potential for both benefit and risk to humanity meets the spirit of the dual use concept (National Research Council, 2004). Given the significant benefits afforded by data sharing and analysis, this paper highlights current data protection policies, potential risks of data exploitation by malicious actors, and potential strategies to mitigate those risks and promote rapid recovery in biotechnology fields that are breached.

The interconnectedness between the digital and biological worlds can be exploited by state actors, malicious nonstate actors, and hackers through a variety of means, resulting in harmful consequences from potential theft of information, promulgation of incorrect information, and/or disruption of activities (Lord and Forbes Technology Council, 2017; Souza, 2018; Ward, 2018). For example, theft of proprietary information from a pharmaceutical or biotechnology company may reveal trade secrets and allow competitors to develop superior products and/or bring existing products to market more quickly (Friedman, 2013), stifling innovation in the global commercial market and allowing adversaries to create harmful, untested therapies. Another example is theft of hundreds of millions of electronic healthcare records, the uses of which are not clear (Bogle, 2018; Cohen, 2018; Healthare IT News Staff, 2018; Huang and Steger, 2018; Keown, 2018). Although unauthorized access to protected data may be aided by technical vulnerabilities in networked computer systems, poor security practices, insider threats in academia, industry, and health facilities, and legal business dealings also can enable adversary access to such data (Lynch, 2017; Rappeport, 2018; South China Morning Post, 2018; Zhu, 2018). For examples, more than half of all data breaches at healthcare facilities are caused by healthcare personnel errors, a quarter of which resulted in unauthorized access to or disclosure of patient records through sharing of unencrypted information, sending information to the wrong patients, and accessing the data without authorization (Bai et al., 2017; Michigan State University, 2018). In addition, the Federal Bureau of Investigation (FBI) has raised national security concerns about foreign access to genomic data of U.S. citizens through legitimate scientific collaboration, funding of scientific research, investment in genomic sequencing companies [e.g., China-based WuXi Healthcare Ventures investment in the U.S.-based 23andMe (Biospace, 2015; Mui, 2016)], and purchase of companies (e.g., Complete Genomics) (Baker, 2012; GenomeWeb, 2012). As vulnerabilities are created through scientific advances, such as the use of machine learning algorithms to trick fingerprint authentication systems, new risks are identified (Bontrager et al., 2018; Nyu Tandon School of Engineering, 2018). Some of these concerns have resulted in the passage of the 2018 Foreign Investment Risk Review Modernization Act, which has initiated reform of the U.S. Government process for evaluating foreign investment in U.S. entities and export control of emerging technologies (Rappeport, 2018; U.S. Congress, 2018). Yet, these policy activities largely are reactive, rather than proactive.

CURRENT APPROACHES FOR PROTECTING DATA

Preventing accidental and deliberate risks typically involves the use of cyber and information security systems that include technological and behavioral solutions. Protection of laboratory control systems, computer networks, and databases often involves the use of technological solutions. However, some risks are addressed better through training of personnel to recognize and report phishing attempts, ensure sensitive information is encrypted, and prevent unauthorized individuals from gaining access to sensitive data, databases, and computer networks. To enhance security, policies for promulgating these practices for specific materials and information have been issued. For example, the U.S. Biological Select Agents and Toxins Regulations include guidance for network security to prevent failure of laboratories, equipment, and access controls to facilities and data (Federal Select Agent Program, 2017). In addition, the U.S. has policies for protecting individual privacy, several of which were described in a 2014 report sponsored by the White House (Podesta et al., 2014). However, error, carelessness, or negligence by personnel can counteract the benefits afforded by security measures and may lead to devastating consequences if biological data and materials are involved.

Although policies for protecting biological data from cyberattack are limited, policies that govern data access and sharing are prevalent. These top-down, data access policies intend to protect individual rights and/or prevent sharing or distribution of data, including biological data. Examples of recent policies include: (a) the 2018 update of the European Union General Data Protection Regulation (European Commission, 2018), which strengthened the European Union's rules for protecting personal data of individuals, in part by giving its citizens "more control over their personal data;" (b) the 2018 Chinese Personal Information Security Specification, which is one system under the Chinese Cybersecurity law, involves the "collection, storage, use, sharing, transfer, and disclosure of personal information," and enables companies operating in China to access data to "not hamper the development of fields like AI" (Sacks, 2018); (c) the 2018 General Data Protection Law in Brazil, which provides a framework for the use of personal data in Brazil (Soares, 2018); and (d) the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA), which promotes the protection of privacy and security of patient health information in the United States (Department of Health and Human Services, 2017). At the same time, the U.S. has issued policies governing data generation, access, and sharing to promote information-sharing and transparency of government-sponsored research (Noorden, 2013). Internationally, the Nagoya Protocol of the Convention on Biodiversity[1] promotes governance on access to and fair, equitable sharing of the benefits from the use of non-human biological data. However, questions exist about whether the Nagoya Protocol focuses more on biological samples that provide genetic information or the genetic information itself, which ultimately affects national-level efforts for codifying the international agreement (Dos et al., 2018). Despite these activities, protection of some data, such as personal health data, may not extend beyond a country's borders and may apply only to data collected by certain entities. Furthermore, data protection polices do not extend to information that already has been stolen. Taken together, these national, regional, and international level policies for data protection may not prevent the inappropriate or unauthorized acquisition of data to different actors, the consequences of which are unclear for biotechnology data.

VULNERABILITY OF BIOTECHNOLOGY DATA

The primary challenges in identifying, assessing, and mitigating security vulnerabilities of biotechnology data are understanding: (a) how the data may be exploited by adversaries and what consequences result from this exploitation; and (2) what potential negative effects may arise from digitalization of biotechnology and advanced computation of biological data (Bajema et al., 2018). The term "biotechnology" refers to the exploitation of biological processes for industrial and scientific purposes, and includes genetic manipulation of microbes, plants, animals, human cells, nucleic acids (the building blocks of genomes), and proteins (the functional units in cells). This definition is expanded further to include generation, incorporation, and use of digital forms of biological data. These biological data may be available online through databases, such as the U.S. National Center for Biotechnology Information's GenBank[2], or generated in a laboratory and stored, shared, and/or analyzed locally or remotely (via online and/or cloud-based software). By attempting to answer the questions posed above, specific risks associated with the legal and illegal acquisition of biological data may be identified and mitigated.

Although extraordinary advances in computing power are enabling unprecedented scientific discoveries, its application to biology and healthcare is increasing without effective protection from the risks of adversary acquisition or accidental misuse of information. Scientific data that is generated in basic and applied research laboratories in academia, non-profit research organizations, service providers, and some industry research facilities may be considered fundamental research destined for publication and public benefit. These data are not necessarily sensitive, but they do represent the results of significant investment by governments, industry, investors, and philanthropic organizations. Therefore, theft or large-scale acquisition of these data may have adverse economic consequences to the organization, field, or nation, especially if acquisition was directed by adversarial nation-states to gain competitive advantage in a given sector (Blair and Huntsman, 2013). As previously described, databases that store sensitive and/or non-sensitive biological data have been infiltrated by external actors and accessed by unauthorized individuals. Although measures to protect data have been implemented in several institutions, cyber and information security policies, practices, and compliance vary across biotechnology sectors, location, and organization type (e.g., academia, industry). Although implementation of cyber, information, and data security in biological facilities can help to minimize the potential for deliberate or accidental release of protected biological data, these measures are insufficient on their own (Press, 2018).

Furthermore, the increasing size and volume of the datasets, and the complexity of analytic technologies has led many scientists to rely on cloud-based platforms to store, transfer, and analyze data. These platforms and technologies, including online analysis software and applications, often do not prevent unauthorized access to data or ensure software fidelity. Although mitigating specific vulnerabilities may be possible on an individual platform or technology level, implementing protections across the various data generation, analysis, transfer, and storage platforms currently in use in academia, industry, government laboratories, and healthcare facilities is challenging. Countering these risks requires the identification of consequences that are of particular concern to public safety

[1] Convention on Biodiversity. About the Nagoya Protocol. Available online at: https://www.cbd.int/abs/about/ (Accessed November 23, 2018).

[2] National Center for Biotechnology Information. GenBank. Available online at: https://www.ncbi.nlm.nih.gov/genbank/ (Accessed November 23, 2018).

and national security, evaluation of vulnerabilities that may enable the realization of these consequences, and identification of measures to address these vulnerabilities.

POSSIBLE PREVENTION AND MITIGATION APPROACHES

Modern cyber and information security reflects the risks experienced as the internet has grown and diversified, and as the capabilities for and speed of storing, processing, and transporting information have increased exponentially (Denning and Lewis, 2017). The internet was built without a priority on the protection of data whether "at rest" (i.e., stored data) or "in motion" (i.e., data in transit) (Dauch et al., 2009; Inap, 2013). Current strategies for addressing cyber risks focus on remediation through regulation, organizational support, and actions taken by data owners and consumers in the form of encryption technologies, access control measures, awareness-raising campaigns, risk assessment, blocking, limiting publication of sensitive information, and other similar practices. The challenge is understanding how these measures are to be applied to biotechnology data, how to balance the cost of implementation with the consequences if left unprotected, and what vulnerabilities cannot be mitigated using commercial products.

Often the entities that assess their cyber vulnerabilities and invest in cyber and information security measures are compelled to do so because of regulation and fiscal responsibility (McDonald, 2017). However, unlike financial information, biotechnology data is regulated in some countries, but not others. For example, China issued a recent policy requiring a domestic collaborator and Ministry-level approval for research involving genomic data of Chinese citizens and/or biological samples obtained in China to prevent exploitation of these data and samples (Tuzman, 2018). This and similar policies raise questions about their intended and unintended effects to nations, to the scientific community, and to international security mainly because the policies that may benefit one country could harm another. These harms may reveal new types of risks associated with the acquisition and use of data to manipulate biological systems. These risks may be perpetrated by different actors; affect sector and country economies, commercial biotechnology, and pharmaceutical markets domestically and internationally; and alter global strategic power dynamics.

The risks associated with biotechnology data do not conform to traditional biosecurity concerns, which focus primarily on risks to human health or the food and agriculture economy. These risks involve multiple domains, sectors, and nations resulting in outcomes such as shifting of balance of power of nations at the international level, which could have downstream effects on areas that overlap with biosecurity interests (e.g., biosafety and biosecurity, biothreat reduction, and global health security). Strategies for bridging the biological, cyber, information, and data security include: (a) collaboration between the biological and cybersecurity communities; (b) end-to-end risk assessments; (c) data-specific risk and vulnerability assessments; and (d) application of the NIST Cybersecurity Framework for protecting biological data.

Formal collaboration between the biotechnology and biological, information, data, and cyber security communities would enhance efforts toward identification of risks and vulnerabilities associated with data management, provenance, and integrity, and risk mitigation strategies. Technologies are readily available to protect data, but their use must be harmonized worldwide, because protecting data in one database is ineffective if another database remains vulnerable to external threats. Furthermore, organizations may evade regulatory requirements and industry standards in protecting data because of perceived lack of cost savings for implementing cybersecurity measures or lack of awareness of the risks, which could lead to investor, intruder, or adversary access to sensitive information that may be stored in databases or transferred between computers. These vulnerabilities may be exacerbated by limitations of national laws to other sovereign states, and differences in interpretation of the types of data included in the scope of existing laws. **Given these potential vulnerabilities, the cybersecurity and biotechnology communities must engage to create best practices and processes to protect data and mitigate risk while reaping the benefits of computing technology applications to biotechnology.**

End-to-end assessments of the data storage, processing, and transport pipeline can identify outstanding vulnerabilities and technical gaps that may be addressed with currently available cyber, information, and data security solutions. This process would enable identification of gaps for which these measures are insufficient and of institutions that are responsible for implementing controls. Without this type of assessment, vulnerabilities may exist along the pipeline without its users' knowledge. A lack of rigorous analysis makes biological data vulnerable to acquisition or alteration by witting adversaries, potentially resulting in theft of intellectual property for commercial gain, foreign government acquisition of genomic data from large portions of a population for undefined purpose or compromise of software and data integrity. At least one country promotes acquisition of data though legitimate commercial practices (e.g., providing sequencing services to customers; partnering with academia, independent research institutions, and universities; and foreign investment), talent promotion programs (Capaccio, 2018; Nature Jobs, 2018), and theft of data (Riley and Walcott, 2015; Dilanian, 2018; Kaiser and Malakoff, 2018; Wilber, 2018). The FBI has expressed concerns about the theft of U.S. genomics and health information through cyberattacks and foreign investment in the U.S. biotechnology industry (You, 2017). The FBI argues that acquisition of this information can give adversaries an unfair advantage in the international pharmaceutical or biotechnology marketplace. Others have expressed concern about questionable use of genetic information that countries obtain from their own citizens or from other countries' citizens (Human Rights Watch, 2017; Lynch, 2017; Pauwels and Vidyarthi, 2017). **These risks could be addressed by conducting an end-to-end risk assessment of the software**

and equipment involved in the data pipeline within individual organizations, between organizations, and across countries.

Defining the consequences of greatest concern to national security is an initial step toward assessing the risks and vulnerabilities of the information itself and data-specific risk mitigation strategies. Evaluating these risks enables the identification of content-specific approaches for detecting and countering exploitation of vulnerabilities by insider and external actors. Without these assessments, only generic cyber and information security measures will be implemented. However, these measures are insufficient to counter adversaries who are intent on acquiring data through a variety of technical, social engineering, or other means. Given this reality, rapid detection and resilience (i.e., rapid recovery after a breach) are critical for reaping the benefits and minimizing the vulnerabilities of advanced electronic computation and mass connectivity. In 2014, the White House explored technology needs for protecting the security and privacy of exposed data, including healthcare data (Executive Office of the President, 2014; President's Council of Advisors on Science Technology, 2014). But, these studies did not define consequences of concern related to the unauthorized acquisition of vast amounts of biological data, effectively limiting the identification of data-specific or process-specific prevention measures. **Therefore, risk assessments of specific types of data are equally as important to conduct as analyses of vulnerabilities of laboratory control systems, data management platforms, and computer networks.**

Application of the National Institute of Standards and Technology (NIST) Cybersecurity Framework to all systems of storage, processing and transport of biological data would help explore where, how, and by whom data is processed with the goal of protecting valuable scientific and health information (National Institute of Standards Technology, 2018). The NIST framework involves a collaboration of private sector and government cybersecurity experts that seek to apply the five principles of data protection (i.e., identify, protect, detect, respond, and recover) to systems, including those on which biological data are generated, processed and transported. The framework could augment existing or newly-implemented efforts of vulnerability detection and mitigation, thus decreasing unauthorized exposure of sensitive data. The NIST framework is a widely accepted paradigm for cyber risk management and best practices (Department of Homeland Security, 2018; Lohrmann, 2018; Roncevich, 2018). In the U.S., this framework has been used in regulatory dialogues to demonstrate rigor toward cybersecurity in sectors for which such requirements are not well-documented in law. **Application of the NIST framework to biotechnology can enhance data protection and a focus on rapid detection of nefarious activity and resiliency after an attack.**

These suggestions describe various approaches toward protecting biological data from unauthorized acquisition and use, enhancing efforts to preserve data integrity and provenance, and enabling future benefit of biotechnological advances.

AUTHOR CONTRIBUTIONS

KB and PS contributed equally to this manuscript. The concepts, conclusions, and recommendations were generated jointly by the authors and built on their respective expertise in the biological sciences and biosecurity, and computer science and cybersecurity.

REFERENCES

Accenture (ed). (2015). *The Future of Applications in Life Sciences: New application Strategies to Unlock the Digital Opportunity*. A.L. Sciences.

Agapakis, C. M. (2014). Designing synthetic biology. *ACS Synth. Biol.* 3, 121–128. doi: 10.1021/sb4001068

Anderson, L. A., Islam, M. A., and Prather, K. L. J. (2018). Synthetic biology strategies for improving microbial synthesis of "green" biopolymers. *J. Biol. Chem.* 293, 5053–5061. doi: 10.1074/jbc.TM117.000368

Bai, G., Jiang, J. X., and Flasher, R. (2017). Hospital risk of data breaches. *JAMA Intern. Med.* 177, 878–880. doi: 10.1001/jamainternmed.2017.0336

Bajema, N. E., Dieuliis, D., Lutes, C., and Lim, Y.-B. (2018). "The digitalization of biology: understanding the new risks and implecations for governance," in *Emergence and Convergence,* ed National Defense University (Washington, DC: National Defense University), 2–3, 7–12.

Baker, M. (2012). China buys US sequencing firm. *Nature* 489, 485–486. doi: 10.1038/489485a

Biospace (2015). *WuXi Healthcare Invests in US Genomics Testmaker 23andMe*. BioSpace. Available online at: https://www.biospace.com/article/releases/-b-wuxi-healthcare-b-invests-in-us-genomics-testmaker-23andme-/

Blair, D. C., and Huntsman, J. M. (2013). *The Report of the Commission on the Theft of American Intellectual Property*. ed T. I. Commission (The National Bureau of Asian Research).

Bogle, A. (2018). *Healthcare Data a Growing Target for Hackers, Cyberseucrity Experts Warn*. ABC News. Available online at: https://www.abc.net.au/news/science/2018-04-18/healthcare-target-for-hackers-experts-warn/9663304 (Accessed November 23, 2018).

Bontrager, P., Roy, A., Togelius, J., Memon, N., and Ross, A. (2018). DeepMasterPrints: generating masterprints for dictionary attacks via latent variable evolution. *arXiv*.

Capaccio, A. (2018). *U.S. Faces 'Unprecidented Threat' From China on Tech Takeover*. Bloomberg. Available online at: https://www.bloomberg.com/news/articles/2018-06-22/china-s-thousand-talents-called-key-in-seizing-u-s-expertise (Accessed November 23, 2018).

Carbonell, P., Jervis, A. J., Robinson, C. J., Yan, C., Dunstan, M., Swainston, N., et al. (2018). An automated design-build-test-learn pipeline for enhanced microbial production of fine chemicals. *Commun. Biol.* 1:66. doi: 10.1038/s42003-018-0076-9

Cha, A. E. (2015). Crowdfunding propels scientific research. *The Washington Post*.

Chapman, T. (2003). Lab automation and robotics: automation on the move. *Nature* 421, 665–666. doi: 10.1038/421665a

Cohen, J. (2018). *Massive Cyberhack by Iran Allegedly Stole Research from 320 Universities, Governments, and Companies*. Science. Available online at: https://www.sciencemag.org/news/2018/03/massive-cyber-hack-iran-allegedly-stole-research-320-universities-governments-and

Dauch, K., Hovak, A., and Nestler, R. (2009). "Information assurance using a defense in-depth strategy," in *Conference For Homeland Security, 2009 CATCH'09, Cybersecurity Applications and Technology* (Washington, DC), 267–272.

Denning, P. J., and Lewis, T. G. (2017). Exponential laws of computing growth. *Commun. ACM.* 60, 54–65. doi: 10.1145/2976758

Department of Health and Human Services (2017). *Summary of the HIPAA Security Rule.* Washington, DC. Available online at: https://www.hhs.gov/

hipaa/for-professionals/security/laws-regulations/index.html (Accessed November 23, 2018).

Department of Homeland Security (2018). *Using the Cybersecurity Framework*. Available online at: https://www.dhs.gov/using-cybersecurity-framework (Accessed January 24, 2019).

Dilanian, K. (2018). *China's Hackers Are Stealing Secrets From U.S. Firms Again, Experts Say*. NBC News. Available online at: https://www.nbcnews.com/news/china/china-s-hackers-are-stealing-secrets-u-s-firms-again-n917836 (Accessed January 27, 2019).

Dos, S. R. C., Koopmans, M. P., and Haringhuizen, G. B. (2018). Threats to timely sharing of pathogen sequence data. *Science* 362, 404–406. doi: 10.1126/science.aau5229

European Commission (2018). *2018 Reform of EU Data Protection Rules*. European Commission.

Executive Office of the President (2014). *Big Data: Seizing Opportunities, Preserving Values*. Washington, DC: White House.

Federal Select Agent Program (2017). *Information Systems Security Controls Guidance.* Atlanta, GA.

Friedman, A. A. (2013). *Cyber Theft of Competitive Data: Asking the Right Questions.* Brookings Institution.

GenomeWeb (2012). *Complete Genomics, BGI Agree to $117.6M Merger*. GenomeWeb. Available online at: https://www.genomeweb.com/sequencing/complete-genomics-bgi-agree-1176m-merger#.XEqIOFxKiUl (Accessed January 24, 2019).

Healthare IT News Staff (2018). *The Biggest Healthcare Data Breaches of 2018 (So Far)*. Healthcare IT News. Available online at: https://www.healthcareitnews.com/projects/biggest-healthcare-data-breaches-2018-so-far (Accessed November 23, 2018).

Huang, E., and Steger, I. (2018). *China is Secretly Enrolling Military Scientists in Western Universities*. Defense One. Available online at: https://www.defenseone.com/threats/2018/10/china-secretly-enrolling-military-scientists-western-universities/152383/?oref=d-mostread (Accessed November 23, 2018).

Human Rights Watch (2017). *China: Minority Region Collects Data from Millions*. New York, NY: Human Rights Watch.

Inap (2013). *Data in Motion vs. Data at Rest.* Available online at: https://www.inap.com/blog/data-in-motion-vs-data-at-rest/ (Accessed January 24, 2019).

Kaiser, J., and Malakoff, D. (2018). NIH investigating whether U.S. scientists are sharing ideas with foreign governments. *Science.* doi: 10.1126/science.aav2343

Keown, A. (2018). *Second Scientist Pleads Guilty to Steeling GlaxoSmithKline Trade Secrets*. BioSpace. Available online at: https://www.biospace.com/article/-jc1n-second-scientist-pleads-guilty-to-stealing-glaxosmithkline-trade-secrets/ (Accessed November 23, 2018).

Lohrmann, D. (2018). *Why You Need the Cybersecurity Framework.* Government Technology.

Lord and Forbes Technology Council (2017). *The Real Threat Of Identity Theft Is In Your Medical Records, Not Credit Cards.* Forbes.

Lynch, D. J. (2017). *Biotechnology: the US-China Dispute over Genentic Data*. Financial Times. Available online at: https://www.ft.com/content/245a7c60-6880-11e7-9a66-93fb352ba1fe (Accessed November 23, 2018).

McDonald, K. (2017). *Private Sector's National Cybersecurity Strategy Contributions Lacking*. TechTarget. Available online at: https://searchcompliance.techtarget.com/opinion/Private-sectors-national-cybersecurity-strategy-contributions-lacking (Accessed January 24, 2019).

Mervis, J. (2017). Data check: U.S. government share of basic research funding falls below 50%. *Science*. doi: 10.1126/science.aal0890

Michigan State University (2018). *Healthcare Providers - Not Hackers - Leak More of Your Data*. EurekAlert!. Available online at: https://eurekalert.org/pub_releases/2018-11/msu-hp-111618.php (Accessed November 23, 2018).

Mui, Y. Q. (2016). *China's $9 Billion Effort to Beat the U.S. in Genetic Testing*. The Washington Post. Available online at: https://www.washingtonpost.com/news/wonk/wp/2016/12/30/chinas-9-billion-effort-to-beat-the-u-s-in-genetic-testing/?noredirect=on&utm_term=.3a83001d622d

National Institute of Standards and Technology (2018). *NIST Cybersecurity Framework.* ed D.O. Commerce. Washington, DC.

National Research Council (2004). *Biotechnology Research in an Age of Terrorism.* Washington, DC: National Academies Press.

Nature Jobs (2018). *China's Plan to Recruit Talented Researchers*. Career Guide. Available online at: https://www.nature.com/articles/d41586-018-\penalty-\@M00538-z

Noorden, R. V. (2013). *White House Announces New US Open-Access Policy*. Nature. Available online at: http://blogs.nature.com/news/2013/02/us-white-house-announces-open-access-policy.html (Accessed November 23, 2018).

Nyu Tandon School of Engineering (2018). *Machine Learning Masters the Fingerprint to Fool Biometric Systems*. PR Newswire. Available online at: https://www.prnewswire.com/news-releases/machine-learning-masters-the-fingerprint-to-fool-biometric-systems-300753375.html

Olena, A. (2018). *Bringing the Internet of Things into the Lab*. The Scientist.

Patel, P. (2018). *DNA Data Storage Gets Random Access*. IEEE Spectrum.

Pauwels, E., and Vidyarthi, A. (2017). "Who will own the secrets in our genes? A U.S.-China race in artificial intelligence and genomics," in *Wilson Briefs,* ed W. W. Center (Washington, DC: Woodrow Wilson Center for International Scholars), 5–9.

Podesta, J., Pritzker, P., Moniz, E. J., Holdren, J., and Zientz, J. (2014). *Big Data: Seizing Opportunities, Preserving Values.* ed E.O.O.T. President (Washington, DC).

President's Council of Advisors on Science and Technology (2014). *Big Data and Privacy: A Technological Perspective.* ed E.O.O.T.U.S. President (Washington, DC: White House).

Press, G. (2018). *60 Cybersecurity Predictions for 2019.* Forbes.

Rappeport, A. (2018). *In New Slap at China, U.S. Expands Power to Block Foreign Investments*. The New York Times. Available online at: https://www.nytimes.com/2018/10/10/business/us-china-investment-cfius.html (Accessed November 23, 2018).

Riley, M., and Walcott, J. (2015). *China's Hack of U.S. Data Tied to Health-Care Record Thefts.* Bloomberg UNE.

Roncevich, T. (2018). *Healthcare IT Security Best Practices: Adopting NIST's Cybersecurity Framework. Cyberguard Compliance*. Available online at: https://info.cgcompliance.com/blog/healthcare-it-security-best-practices-adopting-nists-cybersecurity-framework (Accessed Jan 24, 2019).

Sacks, S. (2018). *China's Emerging Data Privacy System and GDPR*. Washington, DC: Center for Strategic and International Studies.

Service, R. F. (2017). DNA could store all of the world's data in one room. *Science.* doi: 10.1126/science.aal0852

Soares, E. (2018). *Brazil: Personal Data Protection Law Enacted. Global Legal Monitor*. Available online at: https://www.loc.gov/law/foreign-news/article/brazil-personal-data-protection-law-enacted/ (Accessed November 23, 2018).

South China Morning Post (2018). *Chinese Funds Pour US$1.4b into US Biotechnology Firms in the First Three Months of the Year*. South China Morning Post. Available online at: https://www.scmp.com/business/global-economy/article/2142351/chinese-funds-pour-us14b-us-biotechnology-firms-first-three (Accessed November 23, 2018).

Souza, C. (2018). *Lessons for Pharma from the Merck Cyber Attack*. PharmExec.com. Available online at: http://www.pharmexec.com/lessons-pharma-merck-cyber-attack (Accessed January 21, 2019).

Thurow, K., Gode, B., Dingerdissen, U., and Stoll, N. (2004). Laboratory information management systems for life science applications. *Org. Proc. Res. Dev.* 8, 970–982. doi: 10.1021/op040017s

Tuzman, K. T. (2018). *Border Security for China's Genomes*. Biocentury Innovations. Available online at: https://www.biocentury.com/bc-innovations/strategy/2018-10-11/balancing-protection-and-translation-china%E2%80%99s-genomic-data-troves

U.S. Congress (2018). *Foreign Investment Risk Review Modernization Act.*

U. S. Government (2012). *United States Government Policy for Oversight of Life Sciences Dual Use Research of Concern.* Washington, DC.

U. S. Government (2014). *United States Government Policy for Institutional Oversight of Life Sciences Dual Use Research of Concern,* Washington, DC.

Von Krogh, G., Battistini, B., Pachidou, F., and Baschera, P. (2012). The changing face of corporate venturing in biotechnology. *Bioentrepreneur* 30, 911–915. doi: 10.1038/bioe.2012.9

Walpole, J., Papin, J. A., and Peirce, S. M. (2013). Multiscale computational models of complex biological systems. *Annu. Rev. Biomed. Eng.* 15, 137–154. doi: 10.1146/annurev-bioeng-071811-150104

Ward, A. (2018). *ISIS's Use of Social Media Still Poses a Threat to Stability in the Middle East and Africa*. RAND. Available online at: https://www.rand.org/blog/2018/12/isiss-use-of-social-media-still-poses-a-threat-to-stability.html (Accessed January 21, 2019).

Wilber, D. Q. (2018). Chinese hackers charged with stealing data from Navy, JPL and U.S. companies. *LA Times*.

You, E. H. (2017). *Safeguarding the Bioeconomy: U.S. Opportunities and Challenges, Testimony for the U.S.-China Economic and Security Review Commission*. Washington, DC: F.B.O. Investigation.

Zhu, J. (2018). *As China Builds Biotech Sector, Cash Floods U.S. Startups*. Reuters. Available online at: https://www.reuters.com/article/us-biotech-china-investment/as-china-builds-biotech-sector-cash-floods-u-s-startups-idUSKCN1M400G (Accessed November 23, 2018).

Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as representing the views and conclusions or official policies and endorsements, either expressed or implied of Griffin Scientific, Promontory Financial Group or the U.S. Government.

Conflict of Interest Statement: KB was employed by Gryphon Scientific. PS was employed by Promontory Financial Group, which is an IBM Company.

The authors declare that the paper was written in the absence of any commercial or financial relationships that would constitute a conflict of interest.

frontiers
in Bioengineering and Biotechnology

PERSPECTIVE
published: 22 March 2019
doi: 10.3389/fbioe.2019.00051

The National Security Implications of Cyberbiosecurity

*Asha M. George**

Blue Ribbon Study Panel on Biodefense, Washington, DC, United States

The cyber- and biological sciences are converging rapidly, creating benefits, new and advantageous applications, and increasing risks to all nations. The parts of the public and private sectors that should be responsible for cyberbiosecurity are not yet sufficiently organized or supported financially. This article addresses the need to ensure that national security policy: (1) assesses cyberbiological risk and incorporates deterrent and enforcement measures; (2) sets forth clear consequences for those individuals and countries that conduct cyberbiological attacks or otherwise compromise cyberbiosecurity, without imperiling the legitimate sharing of scientific data and information; (3) establishes voluntary cyberbiosecurity standards in partnership with the private sector; (4) identifies cyberbiosecurity threats, vulnerabilities, consequences, and solutions; and (5) results from the combined efforts of all branches of government and the private sector.

Keywords: cyberbio, cyberbiosecurity, cybersecurity, biosecurity, convergence

OPEN ACCESS

Edited by:
Randall Steven Murch,
Virginia Tech, United States

Reviewed by:
Gerald Epstein,
National Defense University,
United States
Calvin Chue,
U.S. Army Edgewood Chemical Biological Center (ECBC),
United States

****Correspondence:***
Asha M. George
asha.george@biodefensestudy.org

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *26 November 2018*
Accepted: *01 March 2019*
Published: *22 March 2019*

Citation:
George AM (2019) The National Security Implications of Cyberbiosecurity.
Front. Bioeng. Biotechnol. 7:51.
doi: 10.3389/fbioe.2019.00051

INTRODUCTION

Many fields of science depend on and are affected by the cyber revolution. The far older field of biology is no exception. In fact, the two fields of biology (the science of life and living organisms, including their physical, chemical, molecular, physiological, and developmental characteristics) and cyberology (the science, study, and theory of cyberspace and cybernetics, including communications over computer networks, Internet-connected systems and data centers, computerized systems, communications and automatic control systems in both machines, and living things) are not only interrelated, each can offer perspectives on the other, enabling greater understanding while simultaneously multiplying the possibilities for new, combined threats, previously unanticipated vulnerabilities, and unintended consequences. Murch et al. (2018) defined cyberbiosecurity as "understanding the vulnerabilities to unwanted surveillance, intrusions, and malicious and harmful activities which can occur within or at the interfaces of comingled life and medical sciences, cyber, cyber-physical, supply chain and infrastructure systems, and developing and instituting measures to prevent, protect against, mitigate, investigate, and attribute such threats as it pertains to security, competitiveness, and resilience." Adequate cyberbiosecurity can only be achieved by taking both cyber- and biological perspectives into consideration simultaneously.

CYBERBIO CONVERGENCE

Lateral thinking intentionally connects disparate subjects to generate new ideas, products, and solutions (de Bono, 1970). Additionally, different scientific areas also converge as we gain greater understanding of their most basic, often elemental characteristics, and comprehend their similarities and sometimes, equivalence (Sharp et al., 2011). Convergence also occurs through the intentional combination of two different fields, using aspects of both to produce something new (Roco and Bainbridge, 2002).

The adjective cyberbio results from all three of these types of convergence. We laterally apply our understanding of biology to robotics, nanotechnology, data, cyberspace, cybernetics, and other cyber-related areas, just as we take our understanding of cyberology and look for the same in biology and biological systems. Organic material developed artificially and used in cyber-enabled technologies and products sometimes behaves in the same way as naturally occurring organic material (Irving, 2017). As we combine the cyber- and biological fields, we create new cyberbio threats, vulnerabilities, and consequences.

National security communities throughout the world cannot afford to ignore cyberbio convergence and the increased requirements for cyberbiosecurity associated with it. As with many scientific advancements, the challenge lies in preventing intended and unintended negative impacts on every nation (Sherden, 2011). Additionally, given the speed at which both cyber- and biological activity can occur independently, the separation between and among nations is already very small. Combined cyberbio activity could move even faster, rendering geographic separation non-existent.

Many critical infrastructure sectors can be affected, and as a result, they must play a role in assuring cyberbiosecurity. The Chemical (particularly due to the convergence of biology and chemistry), Critical Manufacturing, Defense Industrial Base, Emergency Services, Energy, Food and Agriculture, Healthcare and Public Health, and Information Technology Sectors are most affected. While some may be aware of the cyberbiological risk to their sectors, they have not yet determined how best to defend against individual cyber- and biological, let alone combined cyberbiological, risks.

Cyberbio deterrence and enforcement pose challenges for national security policymakers (Blue Ribbon Study Panel on Biodefense., 2015). It is unclear what deterrence measures can be developed or enforced in this regard, especially when deterrence and enforcement are lacking for cyber- and biological activities, individually. With regard to cybersecurity, increased support for overt counter-cyber activities and dedicated cybersecurity agencies (e.g., the governmental mitosis that first resulted in the National Security Agency and U.S. Cyber Command, and then other federal organizations, such as the Department of Homeland Security Cybersecurity and Infrastructure Security Agency, in the United States) may appear to be so large or prolific as to serve as deterrents, but it unclear how effective they will be (Nakashima, 2018). The Biological and Toxin Weapons Convention (Findlay, 2006), programs to control biological select agents (US Government Accountability Office, 2017), and laws and regulations prohibiting the use of biological material for crime, terrorism, and warfare (Hodge, 2012), create some barriers to misuse and establish some agreed upon national and international norms, but serve as imperfect deterrents in the biological arena. Deterrents and laws preventing malevolent cyberbio activity have not been legislated in many countries. Extant legislation addressing cyber- and biological risks lags behind technological advances in these fields and cannot be depended upon to address combined cyberbiological threats, vulnerabilities, and consequences.

CONSEQUENCES WITHOUT IMPERILING LEGITIMATE INFORMATION SHARING

The biological research community depends on digital systems to store and analyze data (Schatz, 2015). Of great concern are the huge amounts of data accessible via the Internet and various Cloud applications, with inadequate cybersecurity (Schneier, 2012). Intellectual property and proprietary information losses associated with digitized biological information could rise to the millions or billions, eventually resulting in economic decreases and reduced international competitiveness (Heus et al., 2017). Other national security concerns include loss of privacy, discrimination, data loss or theft, industrial and commercial sabotage, industrial hacking, exploitation of research to increase disease severity, targeting based on specific DNA patterns, and the production of dangerous and novel pathogens without physical samples (Bajema et al., 2018).

Many of the same countries that are investing large amounts in cutting-edge biological research and dual-use activities that could be used to produce biological weapons are also thought to be responsible for many of the cyber incidents with which the public and private sectors throughout the world struggle today. Advances in cyber- and biological science depend in large part on information systems and management, data storage, and the increased efficiency that computational analysis affords. Some countries may want data and information to feed their growing cyber- and biological weapons programs, increase disease and cyber-attack severity on enemy populations, target specific groups for attack, harm other economies, and boost their own economic competitiveness. Evidence of and information regarding cyberbio convergence and related products may well be the most valuable of all, allowing for the acceleration of nascent, ineffective, or slow-to-develop programs.

While we must encourage the legitimate sharing of scientific data and information, and comprehend that there are not yet reasonable or better alternatives to current cyber communications and data storage options, we must also recognize that all nations and their biological and cyberbiological research, development, science, and technology are at great risk. As a matter of national security, each country must require additional biosecurity and cybersecurity in this arena and set forth clear consequences for individuals and countries who intentionally breech whatever security measures they already utilize to obtain biological and cyberbiological data and information. We must also set forth clear consequences for individuals who do not take enough care to protect the data they generate. Increased cyberbiosecurity may make information sharing more difficult, but it will not make the legitimate sharing of data and information impossible.

ESTABLISHMENT OF VOLUNTARY CYBERBIOSECURITY STANDARDS

The public and private sectors agree with the need for increased cyberbiosecurity. No one is interested in losing their work to their competitors within or outside their organization, company, or country. No one is so naïve as to believe that the nobility of their efforts somehow serves as a protective shield against those who want to further their own agenda.

Considering the vast number of cyber-, biological, and cyberbiological efforts currently underway, and the inability of the private sector to protect itself against all national security threats, national governments should work with their private sectors to establish voluntary standards for cyberbiosecurity. Even if governments possess enough knowledge of the breadth and specificity of private sector research and development, they generally have few mechanisms with which to force the private sector to protect against cyberbiological threats.

There are many models for the development and implementation of standards that both the public and private sectors agree to meet (National Research Council., 2015). Fewer models exist to successfully develop incentives for meeting, and agree upon penalties for not meeting, standards. The government must work with the private sector to develop cyberbiosecurity standards, incentives, and penalties within a specified, relatively short period (e.g., 1 year). The speed at which benevolent and malevolent activity is occurring defies the protracted consensus-driven processes in which many governments, such as that of the United States, engage (The White House., 1998).

IDENTIFICATION OF CYBERBIOLOGICAL RISK AND OTHER SOLUTIONS

While both cybersecurity and biosecurity efforts are underway (with more money and resources currently going to the former), there is an obvious gap when it comes to cyberbiosecurity. For example, even within the U.S. Department of Defense, which now possess two powerful cybersecurity organizational elements (i.e., National Security Agency, U.S. Cyber Command) as well as several organizations that conduct biological research and development using highly dangerous pathogens (e.g., U.S. Army Medical Research Institute of Infectious Diseases), efforts to ensure cyberbiosecurity are insufficient (Knapp, 2018). Governmental agencies throughout the world with responsibilities for agriculture, defense, energy, justice, labor, natural resources, and transportation address cyber- and biological threats separately. Departments of justice and other departments that investigate criminal and terrorism financing are also hobbled by weak or non-existent laws for cyberbiological and other new threats.

Some nations combine their military and intelligence activities. Others are fortunate enough to have enough resources to support both separately. In either case, military and intelligence communities throughout the world must acknowledge ongoing cyberbiological activities. These communities often lack the scientific and technological expertise needed to understand the state of science in the cyber- and biological fields, impact of their convergence, intended outcomes for investments in these areas, and how they could and do impact national security. Given the speed with which advances are occurring, intelligence communities throughout the world must assess cyberbiological capabilities, applications, and abilities to do harm. Military and other national security departments must utilize this intelligence to determine how best to protect national assets.

Each country needs a large-scale program to identify and assess cyberbiological risk. At a minimum, such a program should identify new cyberbio threats, vulnerabilities, and consequences (e.g., those associated with pathogen and biomanufacturing data systems, dual-use synthetic biology, biological intellectual property, bioeconomy). This program should result from a public-private partnership among all government agencies, and private sector companies, academic institutions, and other non-governmental organizations. Risk analysis should be rigorous, independent, critical, and comprehensive, utilizing the same or similar methodologies already developed for systems analysis.

As with all areas which are converging presently, expertise is usually very hard to come by. There are some, however, who have worked in or with both fields, who could serve as effective translators between the cyber- and biological communities. Lateral thinkers, who know how to expertly apply knowledge gained in one area to that of another to come up with new insights can also be effectively utilized. As with all relatively new threats, few experts exist now with operational expertise, but they can be developed through academic and operational training and education programs. Intelligence communities should seek to develop insiders involved in cyberbio activities. Public and private sector organizations that address futures must develop scenarios that are used to develop agricultural, diplomatic, healthcare, public health, and military requirements. Governmental and non-governmental scientists must work together to understand and address the problem, while simultaneously contributing to the cyberbio body of knowledge.

COMBINED GOVERNMENTAL EFFORTS

The legislative bodies and those government agencies responsible for implementing laws must work together to reduce national cyberbiological risk.

Legislative bodies must authorize national cyberbiosecurity programs that:

- Address cyberbiological risk and incorporate deterrent and enforcement measures;
- Set forth clear consequences for individuals or countries that undertake such actions without imperiling the legitimate sharing of scientific data and information;
- Allow for the establishment of voluntary standards in partnership with the private sector;
- Identify new cyberbiosecurity threats, vulnerabilities, and consequences; and
- Develop and implement solutions.

Knowing what a government must authorize is less difficult than determining legislative jurisdiction in the cyberbio arena. It is unrealistic to expect that different elements of legislative bodies that have historically addressed either cyber- or biological risk separately will suddenly or automatically work together to develop and pass legislation that address cyberbiological risk. However, given the extremely large potential impact on each nation's bioeconomy, those legislative elements that address commerce, science, and security are best positioned to produce needed cyberbiological legislation.

Each government should also request funding in, and appropriate funding for, their budget for a national cyberbiosecurity program. Given the present cyberbiological risk to all countries, every national leader should immediately add responsibilities to reduce this risk to already funded cybersecurity and biosecurity programs and assign cyberbiosecurity oversight to a very senior-level dedicated position in their governments (e.g., the U.S. Special Assistant to the President and Senior Director for Weapons of Mass Destruction and Biodefense). Leadership should also require evaluation of cyberbiological risk to their national economies.

CONCLUSION

All countries, including the United States, face risks from many sources. Collective dependence on the Internet and electronic communications, cyber- and biological contributions to national and global economies, competitive participation in the biorevolution, and new types of combinational weapons make the need to reduce cyberbiological risk both imperative and vital. We must take the opportunity afforded to us now to eliminate this transnational security gap, before it is exploited by our enemies.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

REFERENCES

Bajema, N. E., DiEuliis, D., Lutes, C., and Lim, Y. (2018). *The Digitization of Biology: Understanding the New Risks and Implications for Governance*. Available online at: https://wmdcenter.ndu.edu/DesktopModules/ArticleCS/Print.aspx?PortalId=97&ModuleId=44472&Article=1569559

Blue Ribbon Study Panel on Biodefense. (2015). *A National Blueprint for Biodefense: Leadership and Major Reform Needed to Optimize Efforts–Bipartisan Report of the Blue Ribbon Study Panel on Biodefense*. Washington, DC: Blue Ribbon Study Panel on Biodefense. doi: 10.13140/RG.2.1.4407.6240

de Bono, E. (1970). *Lateral Thinking*. NewYork, NY: Harper and Row.

Findlay, T. (2006). *Verification and the BWC: Last Gasp or Signs of Life? Arms Control Today*. Available online at: https://www.armscontrol.org/act/2006_09/BWCVerification

Heus, J. J., de Pauw, E. S., Leloux, M., Morpugo, M., Hamblin, M. R., and Heger, M. (2017). Importance of intellectual property generated by biomedical research at universities and academic hospitals. *J. Clin. Transl. Res.* 3:5. doi: 10.18053/jctres.03.201702.005

Hodge, J. G. (2012). The evolution of law in biopreparedness. *Biosecurity Bioterror.* 10, 38–48. doi: 10.1089/bsp.2011.0094

Irving, M. (2017). *Artificial Evolution Aims to Create Life Out of Non-Living Matter*. New Atlas. Available online at: https://newatlas.com/recreating-evolution-test-tube/48856/

Knapp, B. (2018). *Researchers are Sounding the Alarm on Cyberbiosecurity, 5th Domain*. Available online at: https://www.fifthdomain.com/dod/2018/02/08/researchers-are-sounding-the-alarm-on-cyberbiosecurity/

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

Nakashima, E. (2018). *Pentagon Launches First Cyber Operation to Deter Russian Interference in Midterm Elections*. Washington Post. Available online at: https://www.washingtonpost.com/world/national-security/pentagon-launches-first-cyber-operation-to-deter-russian-interference-in-midterm-elections/2018/10/23/12ec6e7e-d6df-11e8-83a2-d1c3da28d6b6_story.html?utm_term=.fc46e6ec038f

National Research Council. (2015). *Standards, Conformity Assessment, and Trade: Into the 21st Century*. Washington, DC: National Academies Press. doi: 10.17226/4921

Roco, M. C., and Bainbridge, W. S. (2002). Converging technologies for improving human performance: integrating from the nanoscale. *J. Nanopart. Res.* 4, 281–295. doi: 10.1023/A:1021152023349

Schatz, M. C. (2015). Biological data sciences in genome research. *Genome Res.* 25, 1417–1422. doi: 10.1101/gr.191684.115

Schneier, B. (2012). Securing medical research: a cybersecurity point of view. *Science* 336, 1527–1529. doi: 10.1126/science.1224321

Sharp, P. A., Cooney, C. L., Kastner, M. A., Lees, J., Sasisekharan, R., Yaffe, M. B., et al. (2011). The third revolution: the convergence of the life sciences, physical sciences, and engineering. Cambridge, MA: Massachusetts Institute of Technology.

Sherden, W. A. (2011). *Best Laid Plans: The Tyranny of Unintended Consequences and How to Avoid Them*. Santa Barbara, CA: Praeger.

The White House. (1998). *Memorandum for Heads of Executive Departments and Agencies (Circular No. A-119 Revised)*. Washington, DC: The White House.

US Government Accountability Office (2017). *High-Containment Laboratories: Coordinated Actions Needed to Enhance the Select Agent Program's Oversight of Hazardous Pathogens*. Washington, DC: Government Accountability Office.

Conflict of Interest Statement: The author was employed by the Blue Ribbon Study Panel on Biodefense.

frontiers in Bioengineering and Biotechnology

PERSPECTIVE
published: 29 March 2019
doi: 10.3389/fbioe.2019.00061

Assessing Cyberbiosecurity Vulnerabilities and Infrastructure Resilience

Daniel S. Schabacker[1]*, *Leslie-Anne Levy*[2], *Nate J. Evans*[1], *Jennifer M. Fowler*[1] and *Ellen A. Dickey*[1]

[1] Argonne National Laboratory (DOE), Strategic Security Sciences Division, Lemont, IL, United States, [2] Argonne National Laboratory (DOE), Decision and Infrastructure Sciences Division, Lemont, IL, United States

OPEN ACCESS

Edited by:
Diane DiEuliis,
National Defense University,
United States

Reviewed by:
Dana Perkins,
United States Department of Health and Human Services, United States
Gerald Epstein,
National Defense University,
United States

***Correspondence:**
Daniel S. Schabacker
dschabacker@anl.gov

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *26 November 2018*
Accepted: *05 March 2019*
Published: *29 March 2019*

Citation:
Schabacker DS, Levy L-A, Evans NJ, Fowler JM and Dickey EA (2019) Assessing Cyberbiosecurity Vulnerabilities and Infrastructure Resilience. Front. Bioeng. Biotechnol. 7:61. doi: 10.3389/fbioe.2019.00061

The convergence of advances in biotechnology with laboratory automation, access to data, and computational biology has democratized biotechnology and accelerated the development of new therapeutics. However, increased access to biotechnology in the digital age has also introduced additional security concerns and ultimately, spawned the new discipline of cyberbiosecurity, which encompasses cybersecurity, cyber-physical security, and biosecurity considerations. With the emergence of this new discipline comes the need for a logical, repeatable, and shared approach for evaluating facility and system vulnerabilities to cyberbiosecurity threats. In this paper, we outline the foundation of an assessment framework for cyberbiosecurity, accounting for both security and resilience factors in the physical and cyber domains. This is a unique problem set, but despite the complexity of the cyberbiosecurity field in terms of operations and governance, previous experience developing and implementing physical and cyber assessments applicable to a wide spectrum of critical infrastructure sectors provides a validated point of departure for a cyberbiosecurity assessment framework. This approach proposes to integrate existing capabilities and proven methodologies from the infrastructure assessment realm (e.g., decision science, physical security, infrastructure resilience, cybersecurity) with new expertise and requirements in the cyberbiosecurity space (e.g., biotechnology, biomanufacturing, genomics) in order to forge a flexible and defensible approach to identifying and mitigating vulnerabilities. Determining where vulnerabilities reside within cyberbiosecurity business processes can help public and private sector partners create an assessment framework to identify mitigation options for consideration that are both economically and practically viable and ultimately, allow them to manage risk more effectively.

Keywords: cyberbiosecurity, vulnerability, resilience, risk, convergence, emerging, converging, technology

INTRODUCTION

An important initial step in effectively managing risk is developing a comprehensive understanding of vulnerabilities. Stakeholders can then identify economical and practical options to mitigate vulnerabilities. Risk in the biological sciences has been managed through the implementation of standard biosecurity practices, through which vulnerabilities are (a) identified and (b) mitigated through regularly updated training, policies, and enhanced physical security. To prevent unauthorized access to high-consequence biological agents, the U.S. Government (USG) stood up the Federal Select Agent Program (FSAP), which added extensive requirements (e.g., background checks, registration by institutions, increased oversight) for those seeking access to Biological Select Agents and Toxins (BSATs). The BSAT list is based on taxonomic classifications and includes 67 high-consequence biological agents and toxins. Advances in genetic engineering tools (e.g., CRISPR Cas 9 systems) along with the convergence of lab automation, computational biology, and access to publically available genomic databases will dramatically impact the effectiveness of the FSAP as well as other biosecurity policies and practices. It will no longer be necessary to obtain physical samples to exploit a biological agent; access to publically available genomic databases, biofoundries, lab automation, and computational biology enables the design and production of high-consequence biological agents and toxins. These biological agents may be entirely new to nature and unconstrained by taxonomic classification such as the BSAT list (Wintle et al., 2017). This new digital environment in which biological research increasingly takes place must be systematically assessed for vulnerabilities in order to effectively manage evolving risks. The new discipline of cyberbiosecurity, which includes biosecurity, cyber-physical security, and cybersecurity, directly addresses the unique risks associated with biotechnology in an increasingly digital environment (Peccoud et al., 2017; Murch et al., 2018).

In this paper, we outline the foundation of an assessment framework for cyberbiosecurity, accounting for both security and resilience factors in the physical and cyber domains. When implemented, the assessment framework will help partners identify and prioritize vulnerabilities. Importantly, the prioritization of vulnerabilities will result from a defensible, transparent, and reproducible assessment. In conjunction with an understanding of the consequences of disruption, risk mitigation strategies can be developed and considered in return-on-investment (ROI) analyses. ROIs will allow stakeholders to make informed decisions on how best to allocate limited resources for maximum impact.

While biosecurity is one of the three disciplines comprising cyberbiosecurity (e.g., biosecurity, cyber-physical security, and cybersecurity) it is well-established and will not be discussed due to space limitations.

RISK MITIGATION IN THE ERA OF CONVERGING TECHNOLOGIES

Emerging and converging technologies present new risks to security that require new methodologies for risk prioritization and mitigation.

The accelerated pace of technological advancements across nearly all scientific disciplines has been driven largely by the convergence of advancements in scientific disciplines associated with computation, networking, automation, and access to data. Convergence occurs where scientific disciplines or key enabling technologies combine with other disciplines or enabling technologies and promise new or improved capabilities. Convergence is more than the simple combination of different disciplines or technologies. It leads to synergies, adding more value through convergence (Dengg, 2018).

While converging technologies lead to fast and far-reaching improvements, they also create new security challenges and risks. We often try to address new risks with methods that were successful in the past; however, they may not be appropriate for the systemic risks posed by the increasing interconnectivity and complexity associated with converging technologies (Dengg, 2018). Additionally, with highly interconnected systems, the risk from dependencies and interdependencies must be considered. Therefore, we must take a more systemic approach to assessing and mitigating risks resulting from converging technologies.

Emerging and converging technologies have significantly increased the number of vulnerabilities to national security to levels that are untenable for the government and private sector to address in their entirety. They simply do not have the resources required to implement mitigation strategies to address risks with a low probability of occurrence and/or low consequence. Current conversations do not prioritize potential courses of action based on defensible integrated risk assessments that consider both probability and consequence in the context of converging technologies.

CYBERBIOSECURITY

The exploration of life sciences has become increasingly dependent upon internet-connected machinery and devices. Internet-dependent infrastructure is critical to computation and discovery of new avenues of research. The subsequent dependence upon technology and internet-connected devices begs the need to secure this infrastructure. For example, attackers could exploit unsecured networks and remotely manipulate biological material, creating new threats with devastating potential (Murch et al., 2018). Cyberbiosecurity aims to understand and reduce the risks associated with conducting research using advanced technologies in the bioscience field. Science exploration depends increasingly upon cloud services, cyber-physical devices, internet-connected machines, remote databases, and many other cyber-vulnerable technologies. This convergence of science and cybersecurity opens the field to a new threat landscape.

Below are two examples of vulnerabilities that may not be individually identifiable in either a biosecurity or a cybersecurity context but are only apparent when both disciplines are considered.

Bringing together advances in synthetic biology and genetic engineering with machine learning, advanced modeling, metabolic engineering and access to publically available databases containing complete genome sequences of pathogens including virulence factors will enable the design of novel high consequence biological agents completely *in silico*. Minimal laboratory infrastructure and equipment would be required. Moreover, the vast array of publically available open source tools enable execution of these processes by less experienced personnel.

Advances in laboratory automation have enabled tacit knowledge (e.g., hands-on know-how), traditionally requiring years of professional laboratory training, to be codified into executable code controlling automated laboratory equipment. The ability of automated laboratory equipment to reproducibly perform tasks once limited to well-trained laboratorians has been monetized in the form of commercial biological production facilities (e.g., biofoundries). These biofoundries may unwittingly produce components of high consequence biological agents solely from digital information provided by the customer. To request synthesis services, the customer simply goes to the website of the biofoundry and uploads the required biological data (e.g., DNA sequences, amino acid sequences, etc.). To obscure the identity and/or functional properties of the final product several biofoundries can be used, each synthesizing seemingly innocuous products representing only a portion of the final product.

Furthermore, contributions to the exploration of science are built upon the open and sharing nature of samples and knowledge. This inherent openness and trust that exist in the scientific community is ripe for exploitation (Peccoud et al., 2017). In order to thwart attackers and keep data secure, it is paramount that the Confidentiality, Integrity, and Availability (CIA triad) of scientific data is upheld in this digital era. Compromising any of the pillars within the CIA triad could lead to unwanted consequences. For example, attackers could:

- Exploit vulnerable infrastructure and steal proprietary sequences from a biotechnology firm, ruining the *confidentiality* of the stolen intellectual property;
- Manipulate DNA sequences for malicious intent, thereby destroying the *integrity* of a given sample or changing a sample to be something other then what is intended; or
- Degrade systems, compromising the *availability* of cyber-physical devices that are used to perform needed functions.

Ensuring the confidentiality, integrity, and availability of both the physical material and the associated digital information is essential to ensuring the safety and security of scientific advances in bioscience.

UNDERSTANDING KEY TERMS

Defining the key elements of the emerging field of cyberbiosecurity is important to ensuring a common understanding of the relevant technical issues that arise from this new hybrid discipline. It is equally important to define key terms related to risk, particularly for audiences that may not already be familiar with the core concepts relevant to biosecurity; cyber-physical security; and cybersecurity assessments, policies, and practices. An important foundational document in this regard is the *DHS Risk Lexicon*, published in 2010 by the U.S. Department of Homeland Security to level-set terminology across the homeland security enterprise (U.S. Department of Homeland Security (DHS), 2010).

As framed in the *DHS Risk Lexicon*, risk is the potential for an unwanted outcome resulting from an incident, event, or occurrence, as determined by its likelihood and the associated consequences. Evaluating the probability of adversarial attacks is challenging due in part to the lack of historical data in which to ground quantitative estimates, inability to project that future deliberate threats will resemble those of the past and the inherent challenges in evaluating the intent and capability of entities seeking to exploit weaknesses. Thus, risk in the Homeland Security space has been framed as a function of three elements: the threats to which an asset or system is susceptible; the vulnerabilities of the asset or system to the threat; and the potential consequences arising from the degradation of the asset or system. Each of these elements is defined below (U.S. Department of Homeland Security (DHS), 2010).

- Threat: natural or human-caused occurrence, individual, entity, or action that has or indicates the potential to harm life, information, operations, the environment, and/or property.
- Vulnerability: physical feature or operational attribute that renders an entity open to exploitation or susceptible to a given hazard.
- Consequence: the effect of an event, incident, or occurrence. Consequence is commonly deconstructed and measured in four categories: human, economic, mission, and psychological.

When talking about risk, it is also important to define what a hazard is due to its direct correlation and impact on vulnerabilities, threats, and consequences of an asset. A hazard is a natural or man-made source or cause of harm or difficulty. Threats are typically directed at an entity, asset, system, network, or geographic area, while a hazard is a natural or accidental phenomenon that is not driven consciously by an adversary.

Although not typically identified as one of the three core factors driving risk, resilience is an additional consideration that impacts assessments of risk and ensuing strategies for managing it. As a result, it is relevant to understanding ways to evaluate cyberbiosecurity. Resilience is the ability to resist, absorb, recover from or successfully adapt to adversity or a change in conditions (U.S. Department of Homeland Security (DHS), 2010). Resilience features play a role in both the vulnerability and consequence variables in risk. Resilience measures can reduce vulnerability to various threats and hazards through protective measures that improve an organization's ability to resist an event or absorb its effects with minimal impact. Similarly, on the consequence side, resilience measures can enhance an entity's ability to quickly adapt and respond to an incident, as well as to recover and return to normal operations more quickly (U.S. Department of Homeland Security (DHS), 2010; Petit et al., 2013b).

Taking into consideration all of these inputs, organizations can institute defensible, repeatable, and actionable processes to analyze risk and ultimately, to make decisions on how to manage it. Risk management is the process of identifying, analyzing, and communicating risk and then accepting, avoiding, transferring or controlling it to an acceptable level and at an acceptable cost (U.S. Department of Homeland Security (DHS), 2010). Risk management involves knowing the threats and hazards that could potentially impact a given organization, the vulnerabilities that render it susceptible to particular hazards, and the various consequences that might result. **Figure 1** illustrates how these various components combine to drive risk-based decision-making (Petit et al., 2013a).

Cyberbiosecurity is a new field that brings together different disciplines in new ways, triggering a pressing need for new thinking in terms of relevant threats, vulnerabilities, and consequences. Existing approaches used in biosecurity, cyber-physical security, and cybersecurity communities provide important foundational concepts and organizing principles, but they do not adequately capture emergent features related to biological and biomedical systems. Biosecurity, cyber-physical security, and cybersecurity are defined below.

- Biosecurity: describes the protection, control and accountability of biological materials in order to prevent their unauthorized access, loss, theft, misuse, diversion or intentional release.
- Cyber-physical security: addresses the potentially high-consequence dependency between physical systems and the special-purpose computers that control and monitor them.
- Cybersecurity: addresses the risks of computer and network systems used for managing processes and sharing and protecting information.

CONSIDERING DEPENDENCIES AND INTERDEPENDENCIES IN CYBERBIOSECURITY

In addition to the concepts defined in the previous section, another concept that is relevant to understanding risk—including but not limited to the cyberbiosecurity domain—is the notion of how dependencies and interdependencies among and between complex systems impact overall risk. Dependencies and interdependencies are key to how the public and private sector understand, analyze, and manage risk within and across critical infrastructure sectors and other complex systems.

A dependency is a *unidirectional relationship* between two assets, in which the operations of one asset affect the operations of the other. For example, a water treatment plant may depend on an external data source to process its water for potability. An interdependency is a *bidirectional relationship* between two assets, in which the operations of both assets affect each other. For example, the water treatment plant requires communications for its supervisory control and data acquisition (SCADA) system, and, in turn, provides water used by the communications system to cool its equipment. An interdependency is effectively a combination of two dependencies—therefore, understanding an interdependency requires analyses of the one-way dependencies that comprise it (Petit et al., 2015).

Effective analysis of dependencies and interdependencies (whether for critical infrastructure, cyberbiosecurity, or other fields of study) requires some basic frameworks for defining, categorizing, and characterizing key features. For example, since infrastructure systems are constantly interacting with their environment and using inputs to generate outputs, it is important to identify where a dependency or interdependency exists within this activity chain. Upstream dependencies are the products or services provided to one system by an external source that are necessary to support its operations and functions. Internal dependencies involve interactions among internal operations, functions, and missions of the system. Downstream dependencies speak to the consumers or recipients who rely on the system's output and are affected by service disruptions or resource degradation (Petit et al., 2015).

Dependencies and independencies are effectively risk multipliers—they can amplify vulnerabilities and consequences that arise from different threats and hazards. For example, loss of a service such as electric power can potentially affect other infrastructure systems that require power to operate, exacerbating the effects of the original power outage and possibly triggering other unanticipated downstream impacts. The presence of dependencies and interdependencies within the cyberbiosecurity domain make the already complex task of understanding risk that much more complicated, requiring analysts not only to evaluate threat, vulnerability, and consequence factors, but also to characterize relevant dependencies and interdependencies that can render complex systems more susceptible to disruption or exploitation.

FOCUSING ON VULNERABILITY

While the field of cyberbiosecurity is new, community members can leverage extensive knowledge and applications from other fields in order to begin stitching together an overarching framework for understanding relevant threats, vulnerabilities, and consequences from a cyberbiosecurity perspective, whether at a facility, system, or organizational level.

Biolabs need an assessment toolkit that: (1) apply to a wide range of assets and systems across different sectors; (2) produce repeatable, defensible, and actionable results; (3) balance the need for efficiency with the need for detailed data; and (4) build on sound scientific principles, industry standards, and recognized best practices. The approaches above have been used to build and deliver multiple infrastructure assessment tools focused on vulnerability (e.g., Infrastructure Survey Tool (IST), Cyber Infrastructure Survey Tool (Cyber IST), Modified Infrastructure Survey Tool) and are based on the principles of decision analysis, an approach that can be used to manage risk under conditions of uncertainty (Keeney and Raiffa, 1976; Kenney, 1992). When combined with additional analyses that evaluate potential threats and consequences of disruptions or loss, these processes can help biosecurity partners understand their broader risk environment and potential courses of action to mitigate risk.

One example application that could be helpful to the biosecurity community is the IST, which DHS field personnel use to evaluate security and resilience at critical infrastructure facilities nationwide in partnership with infrastructure owners and operators. The IST includes an index—the Protective Measures Index (PMI)—that characterizes the protective measures posture of individual facilities based on their most vulnerable aspects (Fisher et al., 2009; Petit et al., 2011). The PMI

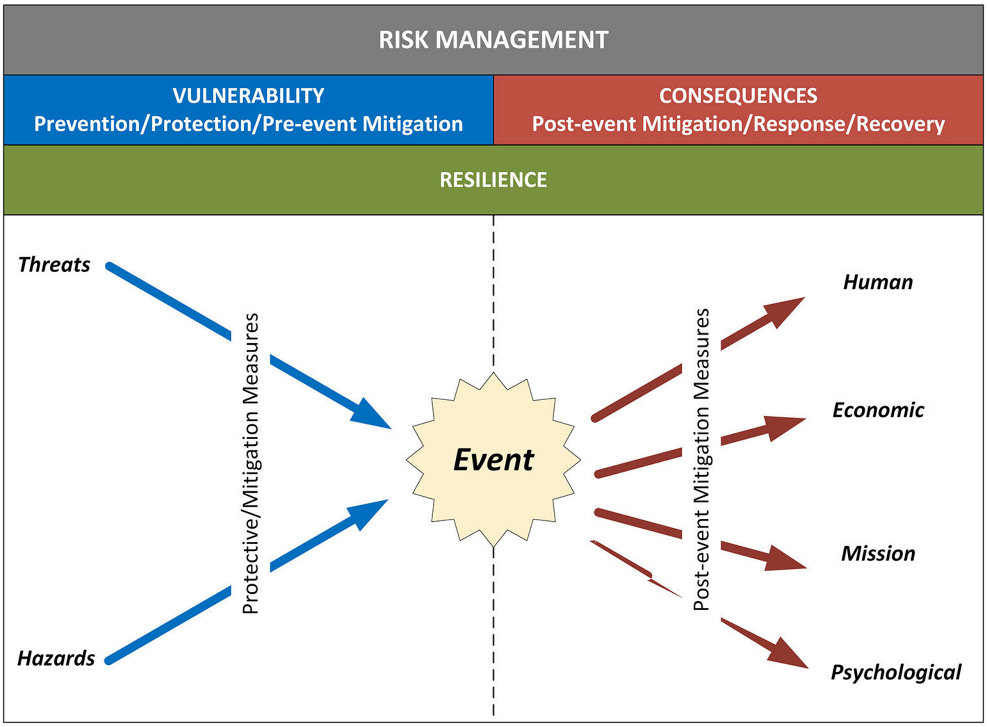

RISK MANAGEMENT
VULNERABILITY
Prevention/Protection/Pre-event Mitigation
CONSEQUENCES
Post-event Mitigation/Response/Recovery
RESILIENCE
Threats
Hazards
Protective/Mitigation Measures
Event
Post-event Mitigation Measures
Human
Economic
Mission
Psychological

FIGURE 1 | Risk management. By understanding the likelihood of various threats and hazards, associated vulnerabilities, potential consequences, and resilience characteristics, stakeholders can make informed decisions on ways to manage risk (i.e., accept, transfer, avoid, or mitigate).

aggregates data collected through a structured onsite assessment process into four levels of information (or subcomponents) across five major categories. For each subcomponent, an index corresponding to the weighted sum of its subcomponents is calculated. This process results in an overall PMI that ranges from 0 (low protection) to 100 (high protection) for the critical infrastructure analyzed, as well as index values for various subcomponents (Petit et al., 2013b).

The decision analysis methodology used to define the PMI was specifically developed to integrate the major elements that are relevant to protecting critical infrastructure. The methodology integrates physical elements that are traditionally part of protection analysis (e.g., fencing, gates, entry controls, intrusion detection systems) as well as operational elements (e.g., security management, security planning, information-sharing mechanisms). The process for identifying specific security characteristics that contribute to protection at a facility and then establishing relative weights required a series of structured elicitation sessions with subject matter experts from public and private sectors (Petit et al., 2013a).

Ultimately, organizing PMI components into different levels and ranking their relative importance allows for the creation of reproducible results and visually compelling outputs that help owners and operators of critical infrastructure make tradeoff decisions on potential courses of action. Furthermore, the use of a consistent index and the consistent deployment of the toolset for a decade has allowed users to compare their results with other assets in the same sector.

Another example that could be helpful to analysis is the Cyber IST, which focuses on critical cyber services. A cyber service is any combination of equipment and devices (hardware), applications and platforms (software), communications, and data that have been integrated to provide specific business services. In this case that would classify as lab systems whose loss would result in physical destruction, safety, and health effects (e.g., a chemical release or loss of environment controls); theft of sensitive information that can be exploited; business interruption (e.g., denial of service); or other economic loss to the organization or its customers/users. The Cyber IST generates a Cyber Protection and Resilience Index (CPRI) as its mechanism for organizations to use in comparative analysis.

In cybersecurity, identified threats, vulnerabilities, and consequences are often categorized into how these risks affect the confidentiality, integrity, and availability of a critical cyber service. These factors are considered the three most significant elements of reliable cybersecurity. Confidentiality limits who has access to information. Integrity governs how and when information is modified. Availability is the assurance that people who are authorized to access the information are able to do so. The question set for the Cyber IST was developed by subject matter experts based on the CIA triad, to assess how businesses help uphold the Confidentiality, Integrity, and Availability of their critical cyber services (Joyce et al., 2017). This same question set provides the basis for assessing confidentiality, integrity, and availability of critical cyber services or assets within the context of cyberbiosecurity.

CONSIDERING THE HUMAN FACTOR IN CYBERBIOSECURITY

Insiders pose substantial threats to cyberbiosecurity because they already have authorized access to critical systems. Most security measures are designed to protect the organization from external attacks and are often more difficult to implement to protect from internal attacks. The potential consequences of threats from insiders vary by the amount of trust and authority given to them (Evans, 2009).

Insiders include not only employees of the organization but also employees of trusted business partners, if those partners have access to the organization's systems, equipment, or data. The threats posed by insiders include both unintentional and intentional, both of which should be accounted for in cyberbiosecurity assessment frameworks. Unintentional incidents often result from negligence or misjudgment. Intentional incidents include insiders who commit fraud for financial gain or seek to sabotage the organization.

Both unintentional and intentional insider incidents can result from actions taken by external actors. For example, unintentional insider incidents may involve insider personnel responding to phishing or social engineering attacks from outside parties, while intentional incidents could involve personnel colluding with external actors, either voluntarily or under pressure. Insiders could willingly participate based on involvement in a cause or support to foreign government or organization, or they may fall victim to recruitment by a criminal enterprise either because of financial or personal troubles (Perkins and Fabregas, 2018).

ROADMAP FOR A CYBERBIOSECURITY ASSESSMENT FRAMEWORK

Moving forward, the diverse community of researchers and practitioners in the cyberbiosecurity domain should collaborate to establish a common vulnerability assessment framework that is grounded in decision science; apply lessons learned from parallel efforts in related fields; and reflect the complex multidisciplinary cyberbiosecurity environment. Key steps in this process should include:

- Engaging subject matter experts in decision science, biotechnology, biosecurity, cyber-physical security, cybersecurity, and physical security in a collaborative assessment development process.
- Defining functional requirements of assessment processes to ensure common understanding of goals, objectives, and constraints.
- Characterizing the biotechnology ecosystem based on facility type (e.g., universities, biofoundries, pharmaceutical companies) and supporting systems (e.g., bioprocess, supply chain, supporting information systems, facility infrastructure) to identify likely assessment candidates and pathways.
- Identifying relevant industry standards, legal frameworks, and regulatory regimes that apply to cyberbiosecurity.
- Establishing a comprehensive taxonomy of characteristics in physical assets and cyber systems in the biotechnology community that influence security posture (e.g., access control, security management, personnel, response protocols, dependencies).
- Conducting an iterative elicitation process to establish subject matter expert consensus on relative importance of security characteristics and their subcomponents in order to facilitate data aggregation, comparison with like entities, and alternatives analysis.
- Exploring potential approaches for collecting assessment data and visualizing assessment results.

AUTHOR CONTRIBUTIONS

DS conceived the manuscript and all authors have jointly contributed to the manuscript, with particular contribution of L-AL to the text on the cyberbiosecurity assessment framework and dependencies. Contribution of NE to the text on cybersecurity and the human factor. All authors have read and approved the manuscript for publication.

FUNDING

Writing of the manuscript was funded by Argonne National Laboratory. Argonne National Laboratory's work was supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357.

REFERENCES

Dengg, A. (Ed.). (2018). *Tomorrow's Technology – A Double-Edged Sword*. Vienna. Available online at: http://www.bundesheer.at/pdf_pool/publikationen/buch_dengg_tomorrows_technology_web.pdf

Evans, N. (2009). *Information Technology Social Engineering: An Academic Definition and Study of Social Engineering-Analyzing The Human Firewall.* Digital Repository@ Iowa State University.

Fisher, R. E., Buehring, W. A., Whitfield, R. G., Bassett, G. W., Dickinson, D. C., Haffenden, R. A., et al. (2009). *Constructing Vulnerability and Protective Measures Indices for the Enhanced Critical Infrastructure Protection Program*, ANL/DIS-09–S-04.

Joyce, A. L., Petit, F. D., Phillips, J. A., Nowak, L. B., and Evans, N. J. (2017). *Cyber Protection and Resilience Index: An Indicator of an Organization's Cyber Protection and Resilience Program.* Available online at: https://www.osti.gov/servlets/purl/1433503

Keeney, R. L., and Raiffa, H. (1976). *Decision Making with Multiple Objectives Preferences and Value Tradeoffs*. New York, NY: Wiley.

Kenney, R. L. (1992). *Value-Focused Thinking: A Path to Creative Decisionmaking*. Cambridge, MA: Harvard University Press.

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol*. 6:39. doi: 10.3389/fbioe.2018.00039

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2017). Cyberbiosecurity: from naive trust to risk awareness. *Trends Biotechnol.* 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012

Perkins, D., and Fabregas, E. (2018). *Mitigating Insider Threats through Strengthening Organizations' Culture of Biosafety, Biosecurity, and Responsible Conduct*. Availiable online at: https://sites.nationalacademies.org/cs/groups/dbassesite/documents/webpage/dbasse_177312.pdf

Petit, F., Bassett, G., Black, R., Buehring, W. A., Collins, M. J., Dickinson, D. C., et al. (2013b). *Resilience Measurement Index: An Indicator of Critical Infrastructure Resilience*. ANL/DIS-13-01.

Petit, F., Bassett, G., Buehring, W. A., Collins, M. J., Dickinson, D. C., Haffenden, R. A., et al. (2013a). *Protective Measures Index and Vulnerability Index: Indicators of Critical Infrastructure Protection and Vulnerability*. ANL/DIS-13/04.

Petit, F., Collins, M., Buehring, W. A., Fisher, R., and Whitfield, R. G. (2011). Protective measures and vulnerability indices for the enhanced critical infrastructure protection program. 7, 200–219. doi: 10.1504/IJCIS.2011.042976

Petit, F., Verner, D., Brannegan, D., Buehring, W. A., Dickinson, D., Guziel, K., et al. (2015). *Analysis of Critical Infrastructure Dependencies and Interdependencies*. ANL/GSS-15/4.

U.S. Department of Homeland Security (DHS) (2010). *Risk Steering Committee. "DHS Risk Lexicon."* Department of Homeland Security. Available online at: https://www.dhs.gov/sites/default/files/publications/dhs-risk-lexicon-2010_0.pdf

Wintle, B. C., Boehm, C. R., Rhodes, C., Molloy, J. C., Millett, P., Adam, L., et al. (2017). A transatlantic perspective on 20 emerging issues in biological engineering. *Elife* 6:e30247. doi: 10.7554/eLife.30247

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Bioengineering and Biotechnology

PERSPECTIVE
published: 29 March 2019
doi: 10.3389/fbioe.2019.00063

Cyberbiosecurity: A New Perspective on Protecting U.S. Food and Agricultural System

Susan E. Duncan[1,2*], *Robert Reinhard*[2,3], *Robert C. Williams*[2], *Ford Ramsey*[4], *Wade Thomason*[5], *Kiho Lee*[6], *Nancy Dudek*[1], *Saied Mostaghimi*[1,7], *Edward Colbert*[8] and *Randall Murch*[4,9]

[1] Virginia Agricultural Experiment Station, Virginia Tech, Blacksburg, VA, United States, [2] Department of Food Science and Technology, Virginia Tech, Blacksburg, VA, United States, [3] Tyson Foods, Chicago, IL, United States, [4] Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, VA, United States, [5] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States, [6] Department of Animal and Poultry Science, Virginia Tech, Blacksburg, VA, United States, [7] Biological Systems Engineering, Virginia Tech, Blacksburg, VA, United States, [8] Hume Center for National Security and Technology, Virginia Tech, Blacksburg, VA, United States, [9] School of Public and International Affairs, Virginia Tech, Arlington, VA, United States

Our national data and infrastructure security issues affecting the "bioeconomy" are evolving rapidly. Simultaneously, the conversation about cyber security of the U.S. food and agricultural system (cyber biosecurity) is incomplete and disjointed. The food and agricultural production sectors influence over 20% of the nation's economy ($6.7T) and 15% of U.S. employment (43.3M jobs). The food and agricultural sectors are immensely diverse and they require advanced technologies and efficiencies that rely on computer technologies, big data, cloud-based data storage, and internet accessibility. There is a *critical need* to safeguard the cyber biosecurity of our bio economy, but currently protections are minimal and do not broadly exist across the food and agricultural system. Using the food safety management Hazard Analysis Critical Control Point system concept as an introductory point of reference, we identify important features in broad food and agricultural production and food systems: dairy, food animals, row crops, fruits and vegetables, and environmental resources (water). This analysis explores the relevant concepts of cyber biosecurity from food production to the end product user (such as the consumer) and considers the integration of diverse transportation, supplier, and retailer networks. We describe common challenges and unique barriers across these systems and recommend solutions to advance the role of cyber biosecurity in the food and agricultural sectors.

Keywords: plant, animal, food, cyber biosecurity, biosecurity, cyber security, agriculture, bio economy

OPEN ACCESS

Edited by:
Stephen Allen Morse,
Centers for Disease Control and Prevention (CDC), United States

Reviewed by:
Jacqueline Fletcher,
Oklahoma State University, United States
Johannes Rath,
University of Vienna, Austria

****Correspondence:***
Susan E. Duncan
duncans@vt.edu

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *11 January 2019*
Accepted: *07 March 2019*
Published: *29 March 2019*

Citation:
Duncan SE, Reinhard R, Williams RC, Ramsey F, Thomason W, Lee K, Dudek N, Mostaghimi S, Colbert E and Murch R (2019) Cyberbiosecurity: A New Perspective on Protecting U.S. Food and Agricultural System. Front. Bioeng. Biotechnol. 7:63. doi: 10.3389/fbioe.2019.00063

INTRODUCTION: FOOD AND AGRICULTURAL CYBERBIOSECURITY AT THE INTERFACE OF BIOSECURITY AND CYBERSECURITY

Public trust and confidence in the food supply are critical and influential on acceptance of data-driven innovations and technologies within the food and agriculture systems (Fd+Ag). Cyberbiosecurity is a nascent paradigm and discipline at the interface of biosafety/biosecurity, cyber security, and cyber-physical security (Murch et al., 2018, **Figure 1**). This new discipline

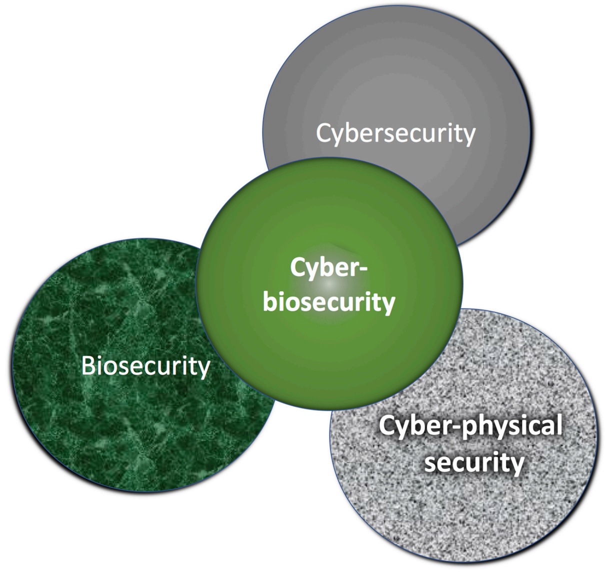


FIGURE 1 | Cyberbiosecurity is an emerging discipline for protecting life sciences data, functions and operations (or infrastructure), and the bio economy.

has emerged alongside "big data" with the extensive and ever-increasing reliance of the life sciences on information systems technologies, rapid and profitable expansion of life science discoveries, and the growth of the U.S. bio economy. Protecting biological data and information within the life sciences has unique differences from the more familiar biosafety and biosecurity approaches (Peccoud et al., 2017). While the latter two categories address biological risks and threats, they do not protect against harm created when computational and information technology-dependent systems are threatened or corrupted. Just as food safety regulations target the protection of human health, incorporating cyber biosecurity strategies for the Fd+Ag system is a protective step in securing the food supply. Such efforts have the power to positively influence lives and protect the bio economy. Cyberbiosecurity can improve the security and stability of the domestic and global Fd+Ag system. Innovation in the U.S. Fd+Ag system is routinely studied and adopted around the globe. The U.S. can provide insight and leadership in cyber biosecurity of the global Fd+Ag systems.

Integrated scientific, mathematical, computational, and engineering advancements in regenerative biology, genetics and breeding technologies, plant-derived vaccine and animal therapies, biological design and testing automation, and other activities are rapidly leading to development of biotechnological and agricultural applications of direct relevance to the Fd+Ag system (The National Academies of Sciences *Engineering and Medicine (NASEM)*, 2014; Wintle et al., 2017). The translation and application of data-driven technologies for precision agriculture, autonomous systems, bio-automated processing and data recording, and other technologies yields large data sets of economic and bio-based information for agribusinesses (Sykuta, 2016). Such advances require high throughput processing, data management and integration, bio-automation, and other computer-based management of biological data. These advances increase efficiencies, decision processes, and output within the food and agricultural system. However, such information is susceptible to ownership policy challenges, theft, and cyber-attack as users may not be alert to potential vulnerabilities nor be trained in effective protections and security strategies (Sykuta, 2016; Boghossian et al., 2018). Unprotected or weakly protected systems are susceptible to unwanted surveillance, intrusions into data systems, and cyber-activities targeted toward malicious attack. Cyberbiosecurity threats include inappropriate access to systems, data, or analytical technologies and the use or corruption of the information accessed to cause harm within life science-focused research, production, processing, and use. Examples of data-driven, high-value food and agricultural products susceptible to cyber threat include high-yielding and specialty agricultural crops, high performance livestock, biopharma fermented molecules developed through advanced breeding and genomics, biotechnology advancements, and "big data" analyses (The National Academies of Sciences Engineering and Medicine, 2015). As technology advances, all parts of society, from governmental agencies to public health and manufacturing, rely more on advanced biological systems with big data and technologies that utilize such information. The identification and mitigation of cyber biosecurity threats will become increasingly important.

VULNERABILITY OF THE FOOD AND AGRICULTURAL SYSTEM AND THE BIOECONOMY

The U.S. Fd+Ag system, influencing 20% ($6.7T) of the domestic bio economy (Feeding the Economy, 2018), represents a significant risk to global food security. The data science market value for agriculture is estimated in excess of $20B (Sykuta, 2016). The Fd+Ag system is composed of many sectors that are not well-integrated, is widely dispersed geographically, and has huge diversity in size (number of employees) and capacity. Most of the economic value in the Fd+Ag system is generated by large, multinational corporate enterprises. Conversely, small family-owned farming operations account for 90% of U.S. farms, which yield 24% of the value of agricultural production (MacDonald and Hoppe, 2017). The family small-business agricultural enterprise (family farm) has economic and social distinctions from corporate farms. Small farm producers view their data with a sense of personal privacy and protection (Sykuta, 2016). Small businesses often use their internet-linked home computer for both personal and business activities, increasing the risk of cyber-attack (United States Department of Agriculture. National Agricultural Statistics Service., 2013; Geil et al., 2018); over 20% of small businesses get hacked (Geil et al., 2018). Generally, small farms and agribusinesses are not comfortable adopting computer security technology (selecting, configuring, managing) although they recognize its relevance and value. Moderate-sized agribusinesses, including many food processing companies and supporting industries, are vulnerable since cyber-attacks are often targeted against organizations with <100 employees (Geil et al., 2018). The Fd+Ag system includes military food production, such as the manufacturing of packaged meals for soldiers, which has a high potential for sabotage (Colbert et al., 2018). It is important to note that attackers need not know details of the food manufacturing process. Attackers need only know technical methods for exploiting the machinery or the process, such as lowering the temperature on meat cookers before packaging (Colbert et al., 2015a,b).

The incorporation of cyber-based technologies and data driven solutions in farm production, food processing, supplier industries, transport of goods, regulatory oversight, and marketing sales and communication with consumers creates a paradigm shift (Boghossian et al., 2018). Cloud-based storage of large data sets, use of open-sourced or internet/cloud-based software, and corporate management of proprietary software each increase opportunities for data access by unauthorized users. Within the Fd+Ag system, the use of biological and genetic analytical technologies within research laboratories is widespread for the evaluation of food quality, identification of zoonotic disease, and animal and plant health. Additionally, the use of bioinformatics and genetic technologies is enhancing the rate of development of new products and crops. Public trust and acceptance are key to incorporating advanced technologies into the Fd+Ag system (United States Department of Agriculture National Institute for Food and Agriculture, 2016; Wintle et al., 2017). Interdependency of information technology with biological output creates opportunities for new bio-threats, which can harm public trust; transparency is valued (The National Academies of Sciences Engineering and Medicine, 2015). When public opinion is turned against a technical advancement, policy and protection strategies may cause more harm than the actual threat itself (Wintle et al., 2017).

Holistically, the ramifications of a failure to provide cyber biosecurity of the Fd+Ag system fall into several general categories (Boghossian et al., 2018):

- Threats to confidentiality—data privacy
    - Data exposure (e.g., naïve exposure of data by individuals, cyber security gaps in small businesses, or laboratories to potential threats);
    - Capturing private data with intent to aggregate data for profit or predictive advantage.
- Threats to integrity—theft or destruction of intellectual property/productivity disruptions, and safety risks
    - Intellectual property theft (e.g., advances in plant and animal varieties and genetics)
    - Manipulation of critical automated (computer-based) processes (e.g., thermal processing time and temperature for food safety);
    - Seizing control of robotics or autonomous vehicles (e.g., failure to perform, overriding of precise function).
- Threats to availability—disruption of agricultural/food production and supply.
- Misinformation influencing trust and cooperation within the Fd+Ag system and/or consumers.
- Lack of equipment, supplies, or end-products to meet expectations;
- Lack of ability to perform vulnerability assessments and develop emergency response plans (e.g., protection of rivers, surface waters, and drinking water supplies).

The food and agricultural industries are at a critical point as the development and use of biological, genetic, precision, and information technologies expand and intersect. Collectively, there is a need to evaluate potential liabilities and understand the vulnerabilities of biological and genetic data systems.

RISK ASSESSMENT, CRITICAL CONTROL POINTS, AND REGULATORY OPTIONS

Cybersecurity risk assessment for industrial control systems (ICS) is advancing rapidly. Cherdantseva et al. (2016) reviewed 24 different cyber security risk assessment methods relevant to ICS. Applications of such risk assessment approaches in Fd+Ag sectors have not been evaluated and the complexity and diversity of the Fd+Ag system may not conform to the current cyber security risk assessment methods. Cyberbiosecurity risk assessment strategies that address the unique security challenges at the intersection of the biological, physical, and cyberspace are important for protecting the Fd+Ag system.

Food manufacturers use the principles of Hazard Analysis and Critical Control Points (HACCP) to assure the production of safe products. HACCP is a familiar risk assessment process within the Fd+Ag system. This management system looks at the likely occurrence of a chemical, biological, or physical food safety hazard in the manufacturing process and the controls that can be put in place to reduce, eliminate, or control the potential hazard. HACCP principles use critical control points (CCPs) as steps in a process where specific controls can be implemented to control, reduce, or eliminate a hazard. HACCP principles are used around the world for the production of safe food products and are required by USDA Food Safety Inspection Service and the U.S. FDA. A risk matrix (**Supplemental Material, Table 1**) may be used to identify potential vulnerabilities and estimate likelihood of occurrence with the potential public health and financial consequences. An example using HACCP principles for an assessment of an Industrial Laboratory processing biological and genetic materials is presented in the **Supplemental Materials**. In this specific example, two CCPs (alternative supplier verification of biological and genetic materials program, and cyber biosecurity data verification program) were identified to mitigate potential risks. Four control point programs (supplier approval; employee training; security programs; and good laboratory standard operating procedures) were identified to support the overarching process for cyber biosecurity.

Several economic problems confront policymakers when addressing cyber biosecurity in the Fd+Ag sector. The most pressing concerns are externalities caused by the networked nature of the system and the misaligned incentives of individual agents. The risks associated with cyber biosecurity threats and harm to society are likely to be larger than the losses suffered by an individual entity; individual firms may not have incentives to provide socially optimal levels of security for the network. Furthermore, if agents know that their own protection depends on security investments made by others, they may become free-riders. Again, this results in inadequate private provision of the public good or security of the network (Varian, 2004).

Multiple regulatory and policy options exist to counter threats to the Fd+Ag system. In some cases, it may be easier to implement protections within the Fd+Ag sector because agribusinesses are already subject to relatively strict disclosure regulations. Information disclosure provides regulators with the data necessary to align individual incentives with the security of the system as a whole. This could be done with top-down regulation, changes to the assignment of liability, or the development of market based systems for the control of cyber biosecurity risks. For instance, the development of cyber biosecurity insurance markets could be encouraged. Regardless of eventual policy measures, it will be important to ensure that the costs of protecting the system are properly aligned with the probabilities of loss and magnitudes of loss associated with cyber biosecurity threats. The most efficient methods of securing the Fd+Ag system are likely to rely on a variety of regulatory approaches.

CONSIDERING THE DIVERSITY WITHIN AND ACROSS PLANT, ANIMAL, AND ENVIRONMENTAL SECTORS OF THE FOOD AND AGRICULTURAL SYSTEM

The HACCP concept assesses risk and establishes CCPs for a specific facility and cannot be generalized effectively to all food manufacturing plants. Applying this concept for cyber biosecurity risk, control points, and CCPs, therefore, is challenged by the diversity of enterprises within a sector and across the Fd+Ag system. Within each sector are unique suppliers providing biological material, chemicals and ingredients, robotics and machinery, software, data, and data storage systems. Some of security measures are encompassed by cyber security, cyberphysical security, and biosecurity/biosafety practices, at least for large corporate entities with sufficient resources. However, an unsecured system from a small agribusiness supplier, producer, processor, or commodity cooperative, could introduce risk.

We use the illustration of a train with multiple boxcars as an example of various sectors within one commodity sector of the Fd+Ag system (**Figure 2**, top). The various cars represent the transition from genetics and breeding through production, processing, distribution, and consumer purchase/use. The exchange of information between the different sectors is often limited, as illustrated by the couplings. The role of the federal government policies and programs provide support and guidance (tracks). Suppliers and other support systems access one or more sectors within a commodity system. The system is driven (engine) by general public (consumers) acceptance of practices and goods, or their fear and mistrust if a risk or threat is perceived. If any stage "derails" or if any supporting agency or organization "buckles" due to a cyber-biosecurity threat or attack, the entire system is at risk, with subsequent risk to the U.S. food supply and the bio economy (**Figure 2**, bottom). Currently, the cyber security industry is not visibly involved in protecting biological data interfacing with the cyber-physical infrastructure supporting the Fd+Ag system.

Some potential mitigations to the issues are possible. Cyberbiosecurity planning and implementation are needed to protect the intellectual and physical (data) property associated with such Fd+Ag priorities. Examples include:

- Plant and animal germplasm, such as old world corn germplasm, microbiology collection (pathogens, fermentation, microbiome) repositories, including economic assessment and protection of data sharing;
- Biocontrolled systems or processes, such as "smart" technology greenhouse data;
- Animal and plant disease diagnostic networks and information sharing;
- Fermentation processing and thermal processing control parameters;
- Freshwater and drinking water supplies and treatment systems.

We further illustrate by outlining some unique considerations for various Fd+Ag commodities.

- Dairy: Selection of genetics for breeding is key to the high milk production in the U.S. dairy industry. Genetic data is highly evaluated as part of the process for breeding. Milk production records are important for establishing high performance animals. While there are some very large dairy herds (>2,000 animals), the U.S. dairy industry is dominated by small to medium farms, many of whom sell their milk through a cooperative structure. Herd health records and drug use are regulated. Data security is variable, and often limited. Fluid milk and dairy food processors do not have detailed records of individual cow production or farm production practices, creating a gap in tracing of information and potential for data breach. Processors utilize computer systems for maintaining processing temperatures, ingredient additions, sanitizing, and cleaning steps.
- Food Animals: Selective breeding is critical to maximize genetic gain during food animal production. For instance, multiple line of breeds are incorporated into swine production to enhance heterogeneity. Pedigree information of the breeds significantly influences selection of founders for the production system. Breach or manipulation of the information can lead to a devastating loss to producers. Recent development in genomic-based selection strategies (Sellner et al, 2007) may also be vulnerable to cyber biosecurity threats as the genomic information can be targeted or exploited. Potential application of genome editing technology in food animals (Telugu et al., 2017) may also generate novel genetic information that could dramatically improve productivity of food animals.
- Row Crops: Similar to the dairy industry, the row crop sector consists of a large number of farms of varying size. Grain

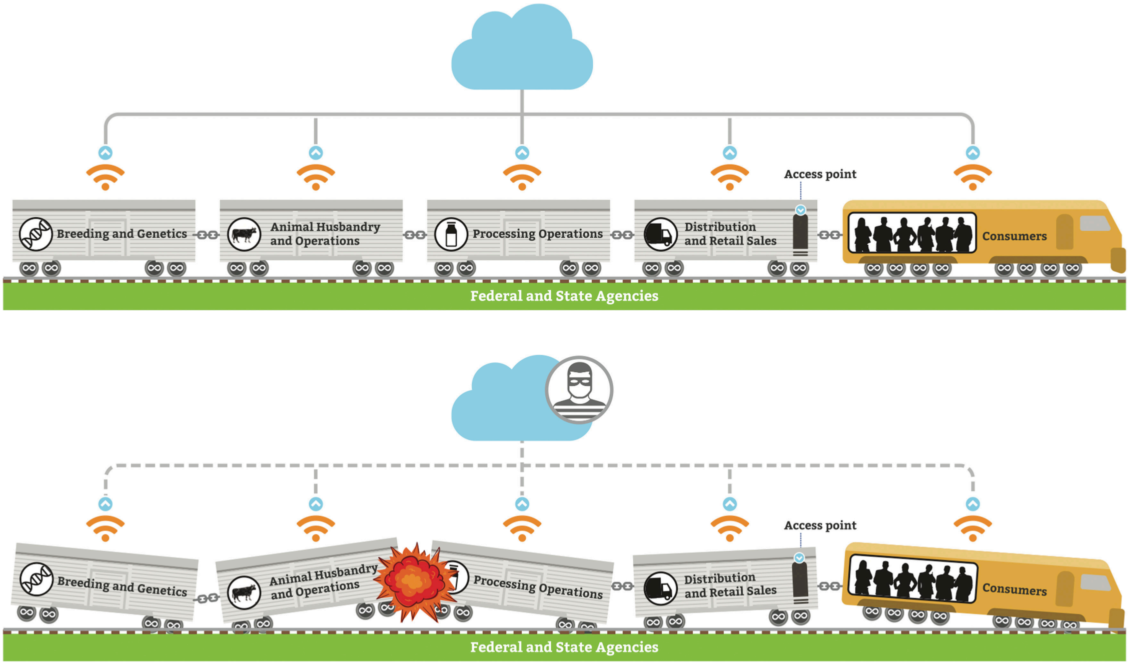

Access point
Breeding and Genetics
Animal Husbandry and Operations
Processing Operations
Distribution and Retail Sales
Consumers
Federal and State Agencies
Access point
Breeding and Genetics
Animal Husbandry and Operations
Processing Operations
Distribution and Retail Sales
Consumers
Federal and State Agencies

FIGURE 2 | (Top) Fd+Ag system for each commodity sector is a sequence of stages, with limited communications and sharing of data between each; **(bottom)** if a cyber-biosecurity event occurs, it can have catastrophic effect on the entire Fd+Ag system.

is typically comingled at the first point of sale and often aggregated further during the process of storage and handling, greatly limiting traceability (Golan et al., 2004). Modern farms using precision agriculture technologies generate enormous amounts of data, about everything from soil conditions to machinery performance and location; such information is often controlled by agriculture technology providers (Sykuta, 2016; Boghossian et al., 2018). Securing data and preventing breaches across all these systems is difficult and is frequently an afterthought by the actual users (Ferris, 2017). Individual producer data is often sent directly to a third party entity for data storage, cleaning, and processing. Many aggregate data and use this as market information or sell it to other companies who do. Commodity traders may use some data streams to guide investment. Anonymization typically occurs at the time of aggregation but questions exist about the effectiveness of these techniques. After transfer, data security becomes the responsibility of the third party data management company, but these entities are themselves not immune from security breaches and would be vulnerable to security issues inserted upstream at the farm or machinery level. Finally, commodity markets are strongly influenced by crop production estimates generated by surveys of farmers and the agriculture industry.

- Fruits and Vegetables: Fresh fruits and vegetables are leading sources for foodborne illness in the United States (Callejón et al., 2015). Furthermore, even in the absence of foodborne illness outbreaks, fresh produce recalls occur regularly due to the presence of potential harmful microorganisms. Fresh produce available for sale in local markets may have been produced in one of many locations throughout the nation or from one of many countries around the world. The production, sorting, grading, commingling, transporting, marketing and sale of fresh fruits and vegetables is complex, and involves numerous industry actors with varying roles. Tracking fresh produce from initial production through consumption is critical to limit the potential for and impact of foodborne illness outbreaks. Accurate product information and rapid access to data is essential to identify contaminated product in the market, prevent or limit foodborne illness, limit the damage to non-implicated producers, and maintain consumer confidence. Access to product tracking and microbiological data is increasing in the fresh produce industry.
- Environmental resources (water): Drinking water safety is extremely important on-farm, for food processing, ensuring the consumers' health and for the proper functioning of the ecosystem. The proportion of the world's population consuming drinking water from certified and controlled water sources is about 90% and still increasing (Vieira, 2011). However, 2.3 billion people worldwide suffer from diseases related to drinking water. Over the past three decades, significant drinking water contamination incidents have occurred in developing as well as developed countries, creating health problems for consumers (Hamilton et al., 2006; Tsoukalas and Tsitsifli, 2018). Traditional risk management systems, based on addressing and correcting the failure after its occurrence, are inadequate to deal with potential cyber biosecurity threats (as the cyber security landscape is changing rapidly as technology continues to advance). Given the severity of risk and potential harm, cyber biosecurity must be given

a high priority for the drinking water management and treatment sector (Germano, 2018).

CONCLUSIONS: MOVING TOWARD SOLUTIONS

The complex and vastly diverse enterprises within the Fd+Ag system increases vulnerability of our food supply and threatens our ability to contribute to the global food supply. Rapid advancements in technologies and adoption into the Fd+Ag sectors increase the risks for cyber biosecurity threats and attacks. The current Fd+Ag workforce has limited knowledge or training appropriate to evaluate and protect the vast amount of data generated by these technologies. The cyber security industry is not well-prepared to address the unique structure and functions within Fd+Ag system. Protecting the Fd+Ag system includes (1) developing and characterizing effective cyber biosecurity risk assessment and mitigation strategies; (2) developing and preparing the current and future workforce to identify, address and adopt effective cyber biosecurity strategies; (3) considering policy and regulations, including insurance, for protection within and across the Fd+Ag system; and (4) effectively communicating within sector and across the Fd+Ag system (United States Department of Agriculture National Institute for Food and Agriculture, 2016). Awareness, knowledge, adoption, and frequent evaluation of cyber biosecurity plans and strategies among and within all Fd+Ag sectors is essential. A multidisciplinary approach integrating expertise in agriculture, food, engineering, computer science, and cyber security is needed for filling this gap. The USDA, in consultation with academic, public and private sector experts and representation from sectors within the Fd+Ag system, should lead an initiative for developing a planned approach to addressing cyber biosecurity. Private and public funding is needed to support research priorities and implementation strategies. Checkoff funding mechanisms or cooperative agreements, which are common within the Fd+Ag commodity systems, may be options for assisting small to moderate-sized agribusinesses. Workforce development, effective communication strategies, and cooperation across sectors and industries will help increase support and compliance, reducing the risks and providing increased protection for the U.S. bio economy and our domestic and global food supply.

AUTHOR CONTRIBUTIONS

SD lead author, responsible for structure, content, and figure; responsible for considering, incorporation co-author contributions and suggested edits; responsible for final version. RR provided draft content related to HACCP and post-harvest processing and cyber biosecurity; contributed, reviewed, and edited the manuscript. RW contributed, reviewed, and edited content related to HACCP and post-harvest processing and biosecurity; reviewed and critiqued manuscript to ensure quality and flow. FR contributed, reviewed, and edited content related to food and agriculture system influence on bio economy; reviewed and critiqued manuscript to ensure quality and flow. WT contributed, reviewed, and edited content related to cyber biosecurity in agriculture (pre-harvest; crop, soil, and environment); reviewed and critiqued manuscript to ensure quality and flow. KL contributed, reviewed, and edited content related to cyber biosecurity in agriculture (pre-harvest; animal breeding and genetics); reviewed and critiqued manuscript to ensure quality and flow. ND contributions to sections relating to biotechnology; overall quality assurance and readability reviews and modifications. SM contributed, reviewed, and edited content related to cyber biosecurity in food and agricultural system and the environment; reviewed and critiqued manuscript to ensure quality and flow. EC contributed, reviewed, and edited content related to cyber security, data sources, and integration into the food and agricultural system; reviewed and critiqued manuscript to ensure quality and flow. RM co-originator of the cyber biosecurity concept; co-originator of the concepts relating to food and agricultural system; contributed, reviewed, and edited content related to cyber biosecurity, data sources, and integration into the food and agricultural system; reviewed and critiqued manuscript to ensure quality, flow, and relevance to the targeted audience.

ACKNOWLEDGMENTS

The authors acknowledge the Virginia Agricultural Experiment Station, Blacksburg, VA for financial support for publishing this manuscript. The authors gratefully acknowledge Michael J. Stamper, Data Visualization Designer and lecturer at the University Libraries, Data Services at Virginia Tech, for his contributions to the figures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2019.00063/full#supplementary-material

REFERENCES

Boghossian, A., Linsky, S., Brown, A., Mutschler, P., Ulicny, B., Barrett, L., et al. (2018). *Threats to Precision Agriculture*. Dept. Homeland Security. Available online at: https://www.dhs.gov/sites/default/files/publications/2018%20AEP__Threats__to__Precision__Agriculture.pdf (Accessed January 08, 2019).

Callejón, R. M., Rodríguez-Naranjo, M. I., Ubeda, C., Hornedo-Ortega, R., Garcia-Parrilla, M. C., and Troncoso, A. M. (2015). Reported foodborne outbreaks due to fresh produce in the United States and European Union: trends and causes. *Foodborne Pathog. Dis.* 12, 32–38. doi: 10.1089/fpd.2014.1821

Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., and Stoddart, K. (2016). A review of cyber security risk assessment methods for SCADA systems. *Comput. Security* 56, 1–27. doi: 10.1016/j.cose.2015.09.009

Colbert, E., Sullivan, D., Wong, K., and Smith, S. (2015a). *Table-Top Exercise Final Report: Intrusion Detection Capabilities for US Army SCADA Systems: Information Packet*. US Army Research Lab Technical Report ARL-TR-7498.

Colbert, E., Sullivan, D., Wong, K., and Smith, S. (2015b). *RED and BLUE Teaming of a US Army SCADA System: Table-Top Exercise Final Report*. US Army Research Lab. Technical Report ARL-TR-7497.

Colbert, E. J. M., Kott, A., and Knachel, L. P. (2018). The game-theoretic model and experimental investigation of cyber wargaming. *J. Defense Model. Simulation*. 1–18. doi: 10.1177/1548512918795061

Feeding the Economy. (2018). *What is the Food and Ag Industries' Impact in your Community*? Available online at: http://feedingtheeconomy.com/ (Accessed October 28, 2018).

Ferris, J. L. (2017). Data privacy and protection in the agriculture industry: is federal regulation necessary. *Minn. JL Sci. Tech*. 18, 309–342.

Geil, A., Sagers, G., Spaulding, A. D., and Wolf, J. R. (2018). Cyber security on the farm: an assessment of cyber security practices in the United States agriculture industry. *Internat. Food Agribus. Manage. Rev*. 21, 317–334. doi: 10.22434/IFAMR2017.0045

Germano, J. H. (2018). *Cybersecurity Risk and Responsibility in the Water Sector*. American Water Works Assn., 20 pages. Available online at: https://www.awwa.org/Portals/0/AWWA/Government/AWWACybersecurityRiskandResponsibility.pdf?ver=2018-12-05-123319-013 (Accessed January 08, 2019).

Golan, E. H., Krissoff, B., Kuchler, F., Calvin, L., Nelson, K., and Price, G. (2004). *Traceability in the US Food Supply: Economic Theory and Industry Studies* (No. 33939). United States Department of Agriculture, Economic Research Service.

Hamilton, P. D., Gale, P., and Pollard, S. J. T. (2006). A commentary on recent water safety initiatives in the context of water utility risk management. *Environ. Internat.* 32, 958–966. doi: 10.1016/j.envint.2006.06.001

MacDonald, J. M., and Hoppe, R. A. (2017). *Large Family Farms Continue to Dominate U.S. Agricultural Production*. USDA ERAmber Waves, S., March 6. Available online at: https://www.ers.usda.gov/amber-waves/2017/march/large-family-farms-continue-to-dominate-us-agricultural-production/ (Accessed October 28, 2018).

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bio economy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2017). Cyberbiosecurity: from naïve trust to risk awareness. *Trends Biotechnol*. 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012

Sellner, E. M., Kim, J. W., McClure, M. C., Taylor, K. H., Schnabel, R. D., Taylor, J. F. (2007). BOARD-INVITED REVIEW: applications of genomic information in livestock. *J. Anim. Sci.* 85, 3148–3158. doi: 10.2527/jas.2007-0291

Sykuta, M. E. (2016). Big data in agriculture: property rights, privacy and competition in ag data services. *Internat. Food Agribusiness Manage. Rev*. 19, 57–74.

Telugu, B. P., Park, K. E., and Park, C. H. (2017). Genome editing and genetic engineering in livestock for advancing agricultural and biomedical applications. *Mamm. Genome Aug*. 28, 338–347. doi: 10.1007/s00335-017-9709-4

The National Academies of Sciences *Engineering and Medicine (NASEM)* (2014). *Meeting Recap, Workshop – Convergence: Safeguarding Technology in the Bioeconomy*. Organized by the Board on Chemical Sciences and Technology and the Board on Life Sciences (Washington, DC).

The National Academies of Sciences Engineering and Medicine (2015). *Meeting Recap, Safeguarding the Bioeconomy: Applications and Implications of Emerging Science.* Organized by Board on Chemical Sciences and Technology (Washington, DC). Available online at: https://www.ehidc.org/sites/default/files/resources/files/Safeguarding%20the%20Bioeconomy_II_Recap%20Final%20090815.pdf (Accessed October 27, 2018).

Tsoukalas, D. S., and Tsitsifli, S. (2018). A critical evaluation of Water Safety Plans (WSPs) and HACCP implementation in water utilities. *Proceedings* 2:600. doi: 10.3390/proceedings2110600

United States Department of Agriculture National Institute for Food and Agriculture (2016). *Results of "Ideas Engine" Stakeholder Input. NIFA Data Summit: Changing the Face, Place, and Space of Agriculture* (Washington, DC). Available online at: https://nifa.usda.gov/sites/default/files/resource/Stakeholder%20Ideas%20Engine%20Input%20-%20Summary%5B1%5D.pdf (Accessed October 28, 2018).

United States Department of Agriculture. National Agricultural Statistics Service. (2013). *Farm Computer Usage and Ownership*. Available online at: http://tinyurl.com/y9pmfee4 (Accessed on October 27, 2018).

Varian, H. (2004). *System Reliability and Free Riding. Economics of Information Security.* Boston, MA: Springer.

Vieira, J. M. P. (2011). A strategic approach for water safety plans implementation in Portugal. *J. Water Health* 9, 107–116. doi: 10.2166/wh.2010.150

Wintle, B. C., Boehm, C. R., Rhodes, C., Molloy, J. C., Millett, P., Adam, L., et al. (2017). A transatlantic perspective on 20 emerging issues in biological engineering. *eLife*. 6:e30247. doi: 10.7554/eLife.30247

Conflict of Interest Statement: RR is employed by Tyson Foods.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Bioengineering and Biotechnology

PERSPECTIVE
published: 05 April 2019
doi: 10.3389/fbioe.2019.00058

Defending Our Public Biological Databases as a Global Critical Infrastructure

Jacob Caswell[1], Jason D. Gans[2], Nicholas Generous[3], Corey M. Hudson[4], Eric Merkley[5], Curtis Johnson[1], Christopher Oehmen[5], Kristin Omberg[5], Emilie Purvine[5], Karen Taylor[5], Christina L. Ting[1], Murray Wolinsky[2] and Gary Xie[2]*

[1] Sandia National Laboratories, Albuquerque, NM, United States, [2] Los Alamos National Laboratory, Bioscience Division, Los Alamos, NM, United States, [3] Los Alamos National Laboratory, Global Security Directorate, Los Alamos, NM, United States, [4] Sandia National Laboratories, Livermore, CA, United States, [5] Pacific Northwest National Laboratory, Richland, WA, United States

OPEN ACCESS

Edited by:
Randall Steven Murch,
Virginia Tech, United States

Reviewed by:
Bruce Budowle,
University of North Texas Health Science Center, United States
Maria José Saavedra,
Universidade de Trás os Montes e Alto Douro, Portugal

***Correspondence:**
Karen Taylor
karen@pnnl.gov

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: 06 January 2019
Accepted: 04 March 2019
Published: 05 April 2019

Citation:
Caswell J, Gans JD, Generous N, Hudson CM, Merkley E, Johnson C, Oehmen C, Omberg K, Purvine E, Taylor K, Ting CL, Wolinsky M and Xie G (2019) Defending Our Public Biological Databases as a Global Critical Infrastructure. Front. Bioeng. Biotechnol. 7:58. doi: 10.3389/fbioe.2019.00058

Progress in modern biology is being driven, in part, by the large amounts of freely available data in public resources such as the International Nucleotide Sequence Database Collaboration (INSDC), the world's primary database of biological sequence (and related) information. INSDC and similar databases have dramatically increased the pace of fundamental biological discovery and enabled a host of innovative therapeutic, diagnostic, and forensic applications. However, as high-value, openly shared resources with a high degree of assumed trust, these repositories share compelling similarities to the early days of the Internet. Consequently, as public biological databases continue to increase in size and importance, we expect that they will face the same threats as undefended cyberspace. There is a unique opportunity, before a significant breach and loss of trust occurs, to ensure they evolve with quality and security as a design philosophy rather than costly "retrofitted" mitigations. This Perspective surveys some potential quality assurance and security weaknesses in existing open genomic and proteomic repositories, describes methods to mitigate the likelihood of both intentional and unintentional errors, and offers recommendations for risk mitigation based on lessons learned from cybersecurity.

Keywords: cyberbiosecurity, biosecurity, cybersecurity, biological databases, machine learning, bioeconomy

INTRODUCTION

Although an openly shared interaction platform confers great value to the biological research community, it may also introduce quality and security risks. Without a system for trusted correction and revision, these shared resources may facilitate widespread dissemination and use of low-quality content, for instance, taxonomically misclassified or erroneous sequences. Furthermore, as these public databases increase in size and importance, they may fall victim to the same security issues and abuses that plague cyberspace to this day. If we act now by developing the databases with quality and security as a design philosophy, we can protect these databases at a much lower cost and with fewer challenges than we currently face with the Internet.

In this Perspective, the authors aim to outline some potential quality assurance and security weaknesses in existing public biological repositories. In section Background: Problems With Public Biological Databases we provide a discussion of errors present in public biological databases and discuss possible security vulnerabilities inherent in their access, publication, and distribution models and systems. Both unintentional and intentional errors are discussed, the latter of which has not been given significant consideration in literature (Moussouni and Berti-Équille, 2013). In section Approaches for Improving Biological Databases, we attempt to introduce greater trust in the data and analyses by providing recommendations to mitigate or account for these errors and vulnerabilities and point to approaches used by other Internet databases. Finally, in section Preliminary Conclusions, we summarize our recommendations.

This Perspective focuses on databases which contain public and freely available data. We recognize that other biological databases exist which contain private, sensitive, or otherwise valuable data (e.g., human genomes). While unauthorized disclosure is not a formal concern in public, non-human databases, safeguarding against intentional or unintentional erroneous content is. Some approaches have been proposed to protect unauthorized disclosure (Kim and Lauter, 2015; Mandal et al., 2018; Ozercan et al., 2018) and, while we don't survey these approaches in this perspective, we note that the public database community may benefit from these ideas as well.

BACKGROUND: PROBLEMS WITH PUBLIC BIOLOGICAL DATABASES

Data Integrity

An important goal for bioinformatics is the continuous improvement of biological databases. Given the rapid nature of this improvement and the rate of data production though, the content of these repositories is not without error. For example, the problem of contaminated sequences has been recognized for nearly two decades, with evidence stating that bacteria and human error are the two most common sources of contamination (Merchant et al., 2014; Strong et al., 2014). Ancient DNA is also particularly affected by human contamination (Pilli et al., 2013). These contaminants are frequently introduced during experiments (Merchant et al., 2014; Ballenghien et al., 2017) from natural associations and insufficient purification (Simion et al., 2017). In the past few years, additional reports have highlighted cases of DNA contamination in published genome data (Witt et al., 2009; Longo et al., 2011), suggesting that DNA contamination may be more widespread than previously thought. We recognize that errors and omissions can occur in open databases both at the sequence and at the metadata levels, but for this Perspective we mainly focus on sequence and taxonomic data concerns for the purposes of illustrating some of the many data integrity challenges possible.

In addition to contaminations, two high profile examples of sequence errors include the reassembly of a misassembled *Francisella tularensis* genome (Puiu and Salzberg, 2008) and the identification of single nucleotide errors in a reference *Tobacco mosaic virus* (TMV) genome (Cooper, 2014). Without a way to flag or remove the erroneous entries, future researchers are left to continually rediscover them. The errors in the *reference* TMV sequence are particularly disturbing. The taxonomic assignment corresponds to a pathogenic strain, but due to two erroneous single nucleotide polymorphisms (SNPs), virions synthesized from the published reference sequence are atypically *not* infectious. Overlooked contaminations in reference genomes can thereby lead to wrong or confusing results and may have major detrimental effects on biological conclusions (Philippe et al., 2011; Laurin-Lemay et al., 2012). While resequencing could be used to identify and correct sequence errors, it is only possible when the original source material is available. For the given example of single nucleotide errors in the TMV genome, the biological sample (sequenced in 1982) no longer exists. In addition to missing samples, samples of high consequence human and agricultural pathogens may not be available for resequencing.

Database integrity considerations for proteomics are generally similar to those for genomics because databases of protein sequences are derived from genome sequencing, via genome annotation and *in silico* translation. A sequence database error is unlikely to result in spurious detection of a protein that is present in the sample (false positive), but it could easily lead to a failure to detect a protein that is present (false negative). This is particularly concerning for discovery of accurate peptide signatures for use in targeted assays, a rapidly growing area of research.

In this section we discussed the issue of errors in genomic and proteomic databases and their impacts for research and application. Sources of these errors may include, among others, entry errors derived from data transfer, original errors derived from source data, and metadata errors (typically provenance-related) derived from the analysis pipeline. Original errors can arise from sequencing and sample preparation instrumentation chemistry, hardware, and software. Metadata errors can arise from bioinformatics software and faulty human interpretation. Each of these errors may be considered noise or the result of some other unintentional cause, but the key problem to note is that each element of the analytical process introduces some level of artifact when creating the analytical product, i.e., what is defined as a peak or a spot, what is the gene scaffold, what is the closed genome, etc. Any difference in process would therefore by its nature have some impact on the final genome. Our goal here is to start drawing connections between these process elements and genome anomalies.

Vulnerabilities and Intentional Tampering

In contrast to the data integrity issues discussed in the prior section, errors may also be *intentionally* introduced into a biological database. For example, consider the hypothetical scenario discussed in Peccoud et al. (2018) whereby a graduate student reads an article and subsequently requests the plasmids described, but receives a faulty sample. It may be that the published sequences were fabricated, or that the source laboratory unwittingly sent faulty plasmids. One could also imagine a scenario where an intentionally mislabeled or harmful sequence is submitted to an open database that could later be unknowingly synthesized in a research setting or, more seriously, in a production capacity. Furthermore, depending on how sequences could be submitted to the database, the adversary may be able to keep the pathogenic sequence from being detected by certain anomaly detection heuristics.

Individuals may also exploit the vulnerabilities inherent in the database as a cyber-system, leading to errors introduced after publication of data despite manipulation and deletion controls. As with any database, biological databases can be compromised, enabling data integrity issues related to insertion, manipulation, exfiltration, and deletion of data, as well as providing a platform for privilege escalation, unauthorized surveillance, or distribution of malware. Ultimately, the effects of the operating environment and the tools used to deliver databases will inform the most appropriate threat model.

APPROACHES FOR IMPROVING BIOLOGICAL DATABASES

In 2000, a workshop titled *Bioinformatics: Converting Data to Knowledge* (National Research Council, 2000) tackled the question of biological database integrity as one of its focus areas. At that time, suggested solutions included building organism-type (e.g., eukaryote) specific grammar-based tools, enabling database self-validation through specialized ontologies, advocating for quality control in laboratories to minimize likelihood of errors, and authorizing only trained curators and annotators to enter data. They also recommend that data provenance be maintained so that the data history and evolution can be understood over time. These approaches fall more-or-less into two categories: ensuring integrity before or during data entry and analyzing data already in a database. Nearly 20 years later, we still emphasize the importance of quality control in laboratories and standardized data entry procedures, but it is clear that errors continue to make their way into databases for a variety of reasons. In this section, we highlight several categories of existing methods to detect data integrity issues in biological databases and outline the strengths and weaknesses of each. We also provide recommendations for improving biological database security.

Automated Approaches for Detecting Anomalies

Some biological databases take the manual curation approach, such as the SwissProt subset of the UniProt (Universal Protein Resource Database). This effort requires significant resources to maintain, consisting of three principal investigators, a large staff and external advisory board (Pundir et al., 2017). Given the complexity and exponential growth of biological data, automatic methods are needed.

Some tools have been developed to assess the technical quality of genome assemblies [e.g., QUAST (Gurevich et al., 2013)], their completeness in terms of gene content [e.g., BUSCO (Simao et al., 2015), ProDeGe (Tennessen et al., 2016)] and even their contamination level [e.g., acdc (Lux et al., 2016), CheckM (Parks et al., 2015)]. Currently there are several analysis pipelines based on various searches to detect potentially contaminated sequences in the published and assembled genome, such as Taxoblast (Dittami and Corre, 2017), homology searches (Kryukov and Imanishi, 2016), GenomePeek (McNair and Edwards, 2015), and a multi-step cleaning process followed by a consensus of rankings (Cornet et al., 2018; Lu and Salzberg, 2018). All these tools require human review or use of additional tools to distinguish true positive from true negative and are therefore not feasible at scale.

Another database quality issue is the automated identification of taxonomically anomalous, questionable, or erroneous GenBank taxonomic assignments. Automated error identification of taxonomic assignments now draws on methods such as anomaly detection, classification, and prediction techniques. These methods have proved impactful in areas like computer vision (Krizhevsky et al., 2012) and natural language processing (Sutskever et al., 2014). They have also been adopted by bioinformatics and computational biology (Larranaga et al., 2006). Much of the work in applying machine learning to biological data is for classification and prediction of metadata, e.g., gene or taxonomy prediction in genomics, and structure and function prediction in proteomics. Verification of sequence metadata contained in a database is then performed by comparing with the predicted metadata from the sequence.

Sequence-based methods to detect taxonomically misclassified bacterial genome sequences tend to be based either on distance measures between pairs of sequences or on consistency with a reference 16S rRNA phylogeny. Common distance metrics include the average nucleotide identity (ANI), digital DNA-DNA hybridization (dDDH), multi-locus sequence analysis (MLSA), k-mer overlap (summarized in Federhen et al., 2016), and information theoretic distances (Li et al., 2004). Given a genome distance, taxonomic misclassifications have been discovered by identifying outlier genomes that exceed a manually determined distance threshold to trusted reference genomes (Goris et al., 2007; Colston et al., 2014; Figueras et al., 2014; Kim et al., 2014; Beaz-Hidalgo et al., 2015; Federhen et al., 2016; Tanizawa et al., 2016). The need for reference genomes is problematic, since approximately 20% of the bacterial genome sequences in GenBank currently (as of August, 2017) do not have a reference (or "type") genome available (NCBI)[1]. The lack of bacterial genomes with a "type" designation is not due to the cost of sequencing, but rather the need to satisfy a specific set of formal requirements (Federhen, 2015), which include submitting culturable isolates to more than one culture collection. This poses a significant challenge for unculturable bacteria.

Distinct from these pairwise distance-based methods, a recent method for identifying taxonomically mislabeled sequences (Kozlov et al., 2016) uses consistency between a given set of taxonomic labels and a phylogenetic tree computed from a multiple sequence alignment of 16S rRNA sequences. This approach uses a single model of evolution to identify sequences whose taxonomic placement is most likely incorrect. However, there are multiple, competing methods for assigning bacterial taxonomy and, in particular, multiple sequence alignment of 16S rRNA can fail to resolve closely related species (Richter and Rossello-Mora, 2009; Kampfer and Glaeser, 2012; Larsen et al., 2014).

Machine learning has been applied to understand the sequences themselves. For example, the tools DeepBind (Alipanahi et al., 2015) and DeepSEA (Zhou and Troyanskaya, 2015) take sequences as input and learn how variations in

[1] NCBI *Bacterial ANI Report* [Online]. Available: ftp://ftp.ncbi.nlm.nih.gov/genomes/ASSEMBLY_REPORTS/ANI_report_bacteria.txt (Accessed).

the sequences can predict function. The successes of these tools coupled with recent research on sequence anomaly detection using long short-term memory (LSTM) recurrent neural networks (RNNs) in cyber security (Brown et al., 2018) could enable a new technique for biological sequence anomaly detection. Finally, if available, machine learning could potentially be applied to data concerning the sequence sources or data submitters themselves to evaluate quality and trustworthiness. However, that discussion is beyond the scope of this Perspective.

Protections Against Intentional Errors

If a trusted method does not exist to ensure the continued quality and revision of content in biological databases, those who use the data should be aware of this risk and account for it in their analysis appropriately. In what follows, we outline previous efforts to develop analytics to detect and mitigate the impact of deliberately introduced database errors, both known and unknown.

Any machine learning analytic is necessarily a product of the data it observes. In an open data environment, an adversary can directly control any subsequent analysis by changing the data to change an algorithm's underlying model (Goodfellow et al., 2014). The focus of a "counter adversarial" approach to data analytics is to harden machine learning methods against the effects of inputs that are designed to mislead supervised (Dalvi et al., 2004; Kantarcioglu et al., 2011; Biggio et al., 2013a) and unsupervised (Dutrisac and Skillicorn, 2008; Biggio et al., 2013b) algorithms. It has been shown that there exist label tampering attacks which significantly decrease the accuracy of a classifier, while being nearly undetectable by standard cross validation tests (Kegelmeyer et al., 2015). In other words, the defender does not know the performance of the classifier has been corrupted. To protect against label tampering, an "ensembles of outlier measures" (EOM) method has been proposed to identify label tampering. The approach relies on a set of attributes that capture the "outlierness" of a sample to predict whether a sample has been tampered with. Tampered samples can then be remediated by changing the sample class label. In the context of a biological database, these labels may be metadata attributes associated with an entry. In the unsupervised machine learning scenario, an adversary may try to subvert a clustering algorithm by, for example, heuristically inserting data points to arbitrarily poison (i.e., merge) (Biggio et al., 2013b) clusters. In the context of a genomics database, poisoning of clusters would significantly reduce the ability to detect anomalous genomic sequences. Kegelmeyer et al. demonstrate that their remediation methodology based on an EOM applies equally well in the unsupervised context (Kegelmeyer et al., 2015).

As vulnerable cyber systems, best cyber practices can also be leveraged to protect biological databases. However, in the context of intentional manipulation of biological databases, special consideration must be given to the ability of these databases to enable production of dangerous biological material. The International Gene Synthesis Consortium (IGSC), for instance, provides two principal protections against the manufacture of malicious genetic material—known as the Harmonized Screening Protocol (International Gene Synthesis Consortium, 2017). The first, is a customer screening. The second is a screening of DNA sequences against a Regulated Pathogen Database (RPD). This database is built from data on the US Select Agent List, the Australian Group List, and other national lists of regulated pathogens. Members of the IGSC agree to translate each synthetic gene into amino acid sequence and test for homology. These are then accepted, reviewed or rejected. The RPD is updated annually and provided to members.

The Harmonized Screening Protocol requires at least two difficult processes—(1) sharing the database and (2) updating the database. Sharing the database requires the maintenance of authentication. Providers and users are part of a shared environment where they need to trust that everyone has an authentic and up-to-date version of the database. Updating the database requires maintenance to avoid "alert fatigue" from false positives and the dangerous potential case of false negatives resulting in malicious manufacture. Maintaining the security of this requires an environment of authentication and active database inspection and curation. For the former, there may be opportunities to incorporate advanced encryption and authentication algorithms being considered in the cyber domain such as blockchain. However, significant computational resource costs must be contended with.

PRELIMINARY CONCLUSIONS

This survey of concerns with biological databases and methods for ensuring database integrity is certainly not exhaustive but represents broad capabilities within data science and cybersecurity today that have shown promise either within computational biology already, or in tackling similar problems in other domains. A goal of the authors is to illuminate these concerns for a wide audience in the context of the historical lessons learned in cyberspace. In the early days of the Internet, the emphasis was on functionality and enabling the actions of largely well-intentioned communities of users. This functionality pervaded every element of our critical infrastructure. However, the same fabric that supports this infrastructure also represents a significant risk. Mitigating this risk after the wide penetration of open functionality is much more difficult than it might have been if the Internet had been created with integrity and security in mind. As biological data becomes a bedrock critical infrastructure for the entire bioeconomy and follows the same exponential trends of size, pervasiveness, and importance as the Internet, we have a unique opportunity to ensure that this capability mitigates current and future risks from a worldwide set of actors. This paper calls out several existing research areas that can be leveraged to protect against accidental and intentional modifications and misuse of public biological databases.

AUTHOR CONTRIBUTIONS

JC, JG, CH, EM, CO, EP, CT, and GX wrote sections of the manuscript. JC, JG, CH, EM, CJ, CO, KO, EP, KT, CT, and GX contributed to the formulation of the perspectives therein. NG, KT, and MW worked on document conception and assembly. All authors contributed to manuscript revision and approved of the submitted content.

FUNDING

This research was supported in part by the Department of Homeland Security Science and Technology Directorate, award HSHQPM-14-X-00069.

ACKNOWLEDGMENTS

The US Department of Energy (DOE) Laboratories are managed by contracted private entities as Federally Funded Research and Development Companies, and as such each laboratory is operated on behalf of DOE, independently from the operating company. Los Alamos National Laboratory is operated for DOE's National Nuclear Security Administration by Triad National Security, LLC under Contract 89233218CNA000001. Pacific Northwest National Laboratory is operated for DOE by Battelle Memorial Institute under contract DE-AC05-76RL01830. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for DOE's National Nuclear Security Administration under contract DE-NA0003525.

Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

Alipanahi, B., Delong, A., Weirauch, M. T., and Frey, B. J. (2015). Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. *Nat. Biotechnol.* 33, 831–838. doi: 10.1038/nbt.3300

Ballenghien, M., Faivre, N., and Galtier, N. (2017). Patterns of cross-contamination in a multispecies population genomic project: detection, quantification, impact, and solutions. *BMC Biol.* 15:25. doi: 10.1186/s12915-017-0366-6

Beaz-Hidalgo, R., Hossain, M. J., Liles, M. R., and Figueras, M. J. (2015). Strategies to avoid wrongly labelled genomes using as example the detected wrong taxonomic affiliation for aeromonas genomes in the GenBank database. *PLoS ONE.* 10:e0115813. doi: 10.1371/journal.pone.0115813

Biggio, B., Corona, I., Maiorca, D., Nelson, B., Šrndić, N., Laskov, P., et al. (2013a). *Evasion Attacks Against Machine Learning at Test Time*. Berlin: Springer, 387–402.

Biggio, B., Pillai, I., Bulò, S. R., Ariu, D., Pelillo, M., and Roli, F. (2013b). "Is data clustering in adversarial settings secure?," in *Proceedings of the 2013 ACM Workshop on Artificial Intelligence and Security*. (Berlin: ACM).

Brown, A., Tuor, A., Hutchinson, B., and Nichols, N. (2018). "Recurrent neural network attention mechanisms for interpretable system log anomaly detection,“ in *Proceedings of the First Workshop on Machine Learning for Computing Systems* (Tempe, AZ: ACM).

Colston, S. M., Fullmer, M. S., Beka, L., Lamy, B., Gogarten, J. P., and Graf, J. (2014). Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using Aeromonas as a test case. *MBio* 5:e02136. doi: 10.1128/mBio.02136-14

Cooper, B. (2014). Proof by synthesis of *Tobacco mosaic virus*. *Genome Biol.* 15:R67. doi: 10.1186/gb-2014-15-5-r67

Cornet, L., Meunier, L., Van Vlierberghe, M., Leonard, R. R., Durieu, B., Lara, Y., et al. (2018). Consensus assessment of the contamination level of publicly available cyanobacterial genomes. *PLoS ONE.* 13:e0200323. doi: 10.1371/journal.pone.0200323

Dalvi, N., Domingos, P., Mausam, S., and Verma, D. (2004). "Adversarial classification," in *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. (Seattle, WA: ACM).

Dittami, S. M., and Corre, E. (2017). Detection of bacterial contaminants and hybrid sequences in the genome of the kelp Saccharina japonica using Taxoblast. *PeerJ.* 5:e4073. doi: 10.7717/peerj.4073

Dutrisac, J. G., and Skillicorn, D. B. (2008). "Hiding clusters in adversarial settings,“ in *2008 IEEE International Conference on Intelligence and Security Informatics* (Kingston, ON), 185–187.

Federhen, S. (2015). Type material in the NCBI Taxonomy Database. *Nucleic Acids Res.* 43(Database issue), D1086–D1098. doi: 10.1093/nar/gku1127

Federhen, S., Rossello-Mora, R., Klenk, H.-P., Tindall, B. J., Konstantinidis, K. T., Whitman, W. B., et al. (2016). Meeting report: GenBank microbial genomic taxonomy workshop (12–13 May, 2015). *Stand. Genomic Sci.* 11:15. doi: 10.1186/s40793-016-0134-1

Figueras, M. J., Beaz-Hidalgo, R., Hossain, M. J., and Liles, M. R. (2014). Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc.* 2, 1–2. doi: 10.1128/genomeA.00927-14

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2014). "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. (Montreal, QC: MIT Press).

Goris, J., Konstantinidis, K. T., Klappenbach, J. A., Coenye, T., Vandamme, P., and Tiedje, J. M. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* 57(Pt 1), 81–91. doi: 10.1099/ijs.0.64483-0

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086

International Gene Synthesis Consortium (2017). "Harmonized Screening Protocol v2.0,“ in *Gene Sequence and Customer Screening to Promote Biosecurity*. International Gene Synthesis Consortium.

Kampfer, P., and Glaeser, S. P. (2012). Prokaryotic taxonomy in the sequencing era – the polyphasic approach revisited. *Environ. Microbiol.* 14, 291–317. doi: 10.1111/j.1462-2920.2011.02615.x

Kantarcioglu, M., Xi, B., and Clifton, C. (2011). Classifier evaluation and attribute selection against active adversaries. *Data Min. Knowl. Discov.* 22, 291–335. doi: 10.1007/s10618-010-0197-3

Kegelmeyer, P., Shead, T. M., Crussell, J., Rodhouse, K., Robinson, D., Johnson, C., et al. (2015). *Counter Adversarial Data Analytics in Sandia.* Technical Report, Sandia National Laboratory.

Kim, M., and Lauter, K. (2015). Private genome analysis through homomorphic encryption. *BMC Med. Inform. Decis. Mak.* 15 (Suppl 5):S3. doi: 10.1186/1472-6947-15-S5-S3

Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64(Pt 2), 346–351. doi: 10.1099/ijs.0.059774-0

Kozlov, A. M., Zhang, J., Yilmaz, P., Glockner, F. O., and Stamatakis, A. (2016). Phylogeny-aware identification and correction of taxonomically mislabeled sequences. *Nucleic Acids Res.* 44, 5022–5033. doi: 10.1093/nar/gkw396

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). "ImageNet classification with deep convolutional neural networks,“ in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1* (Lake Tahoe: Curran Associates Inc.).

Kryukov, K., and Imanishi, T. (2016). Human contamination in public genome assemblies. *PLoS ONE.* 11:e0162424. doi: 10.1371/journal.pone.0162424

Larranaga, P., Calvo, B., Santana, R., Bielza, C., Galdiano, J., Inza, I., et al. (2006). Machine learning in bioinformatics. *Brief Bioinformatics* 7, 86–112. doi: 10.1093/bib/bbk007

Larsen, M. V., Cosentino, S., Lukjancenko, O., Saputra, D., Rasmussen, S., Hasman, H., et al. (2014). Benchmarking of methods for genomic taxonomy. *J. Clin. Microbiol.* 52, 1529–1539. doi: 10.1128/JCM.02981-13

Laurin-Lemay, S., Brinkmann, H., and Philippe, H. (2012). Origin of land plants revisited in the light of sequence contamination and missing data. *Curr. Biol.* 22, R593–R594. doi: 10.1016/j.cub.2012.06.013

Li, M., Chen, X., Li, X., Ma, B., and Vitanyi, P. M. B. (2004). The similarity metric. *IEEE Trans. Inf. Theor.* 50, 3250–3264. doi: 10.1109/TIT.2004.838101

Longo, M. S., O'Neill, M. J., and O'Neill, R. J. (2011). Abundant human DNA contamination identified in non-primate genome databases. *PLoS ONE*. 6:e16410. doi: 10.1371/journal.pone.0016410

Lu, J., and Salzberg, S. L. (2018). Removing contaminants from databases of draft genomes. *PLoS Comput. Biol.* 14:e1006277. doi: 10.1371/journal.pcbi.1006277

Lux, M., Kruger, J., Rinke, C., Maus, I., Schluter, A., Woyke, T., et al. (2016). acdc - Automated contamination detection and confidence estimation for single-cell genome data. *BMC Bioinformatics* 17:543. doi: 10.1186/s12859-016-1397-7

Mandal, A., Mitchell, J. C., Montgomery, H. W., and Roy, A. (2018). "Data oblivious genome variants search on Intel SGX," in *IACR Cryptology ePrint Archive*. Available online at: https://eprint.iacr.org/eprint-bin/cite.pl?entry=2018/732

McNair, K., and Edwards, R. A. (2015). GenomePeek-an online tool for prokaryotic genome and metagenome analysis. *PeerJ* 3:e1025. doi: 10.7717/peerj.1025

Merchant, S., Wood, D. E., and Salzberg, S. L. (2014). Unexpected cross-species contamination in genome sequencing projects. *PeerJ* 2:e675. doi: 10.7717/peerj.675

Moussouni, F., and Berti-Équille, L. (2013). "Cleaning, integrating, and warehousing genomic data from biomedical resources," in *Biological Knowledge Discovery Handbook*, eds M. Elloumi and A. Y. Zomaya (Hoboken, NJ: John Wiley and Sons, Inc.), 35–58.

National Research Council (2000). *Bioinformatics: Converting Data to Knowledge: Workshop Summary.* Washington, DC: The National Academies Press.

Ozercan, H. I., Ileri, A. M., Ayday, E., and Alkan, C. (2018). Realizing the potential of blockchain technologies in genomics. *Genome Res.* 28, 1255–1263. doi: 10.1101/gr.207464.116

Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2018). Cyberbiosecurity: from naive trust to risk awareness. *Trends Biotechnol.* 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T., Manuel, M., Worheide, G., et al. (2011). Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biol.* 9:e1000602. doi: 10.1371/journal.pbio.1000602

Pilli, E., Modi, A., Serpico, C., Achilli, A., Lancioni, H., Lippi, B., et al. (2013). Monitoring DNA contamination in handled vs. directly excavated ancient human skeletal remains. *PLoS ONE*. 8:e52524. doi: 10.1371/journal.pone.0052524

Puiu, D., and Salzberg, S. L. (2008). Re-assembly of the genome of *Francisella tularensis*. Subsp. holarctica OSU18. *PLoS ONE*. 3:e3427. doi: 10.1371/journal.pone.0003427

Pundir, S., Martin, M. J., and O'Donovan, C. (2017). UniProt protein knowledgebase. *Methods Mol. Biol.* 1558, 41–55. doi: 10.1007/978-1-4939-6783-4_2

Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U.S.A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. doi: 10.1093/bioinformatics/btv351

Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D. J., Di Franco, A., et al. (2017). A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr. Biol.* 27, 958–967. doi: 10.1016/j.cub.2017.02.031

Strong, M. J., Xu, G., Morici, L., Splinter Bon-Durant, S., Baddoo, M., Lin, Z., et al. (2014). Microbial contamination in next generation sequencing: implications for sequence-based analysis of clinical samples. *PLoS Pathog.* 10:e1004437. doi: 10.1371/journal.ppat.1004437

Sutskever, I., Vinyals, O., and Le, Q. V. (2014). "Sequence to sequence learning with neural networks," in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2* (Montreal, QC: MIT Press).

Tanizawa, Y., Fujisawa, T., Kaminuma, E., Nakamura, Y., and Arita, M. (2016). DFAST and DAGA: web-based integrated genome annotation tools and resources. *Biosci. Microbiota Food Health* 35, 173–184. doi: 10.12938/bmfh.16-003

Tennessen, K., Andersen, E., Clingenpeel, S., Rinke, C., Lundberg, D. S., Han, J., et al. (2016). ProDeGe: a computational protocol for fully automated decontamination of genomes. *ISME J.* 10, 269–272. doi: 10.1038/ismej.2015.100

Witt, N., Rodger, G., Vandesompele, J., Benes, V., Zumla, A., Rook, G. A., et al. (2009). An assessment of air as a source of DNA contamination encountered when performing PCR. *J. Biomol. Tech.* 20, 236–240.

Zhou, J., and Troyanskaya, O. G. (2015). Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* 12, 931–934. doi: 10.1038/nmeth.3547

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Bioengineering and Biotechnology

REVIEW
published: 15 May 2019
doi: 10.3389/fbioe.2019.00106

Cyberbiosecurity Challenges of Pathogen Genome Databases

Boris A. Vinatzer[1]*, Lenwood S. Heath*[2]*, Hussain M. J. Almohri*[3]*, Michael J. Stulberg*[4]*, Christopher Lowe*[5] *and Song Li*[1]*

[1] School of Plant and Environmental Sciences, College of Agriculture and Life Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, *[2] Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States,* *[3] Department of Computer Science, Kuwait University, Kuwait City, Kuwait,* *[4] Animal and Plant Health Inspection Service (USDA), Riverdale Park, MD, United States,* *[5] Beltsville Agricultural Research Center, Agricultural Research Service (USDA), Beltsville, MD, United States*

OPEN ACCESS

Edited by:
Diane DiEuliis,
National Defense University,
United States

Reviewed by:
Maria José Saavedra,
Universidade de Trás os Montes e Alto Douro, Portugal
Masfique Mehedi,
University of North Dakota,
United States

****Correspondence:***
Song Li
songli@vt.edu

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *21 February 2019*
Accepted: *25 April 2019*
Published: *15 May 2019*

Citation:
Vinatzer BA, Heath LS, Almohri HMJ, Stulberg MJ, Lowe C and Li S (2019) Cyberbiosecurity Challenges of Pathogen Genome Databases. Front. Bioeng. Biotechnol. 7:106. doi: 10.3389/fbioe.2019.00106

Pathogen detection, identification, and tracking is shifting from non-molecular methods, DNA fingerprinting methods, and single gene methods to methods relying on whole genomes. Viral Ebola and influenza genome data are being used for real-time tracking, while food-borne bacterial pathogen outbreaks and hospital outbreaks are investigated using whole genomes in the UK, Canada, the USA and the other countries. Also, plant pathogen genomes are starting to be used to investigate plant disease epidemics such as the wheat blast outbreak in Bangladesh. While these genome-based approaches provide never-seen advantages over all previous approaches with regard to public health and biosecurity, they also come with new vulnerabilities and risks with regard to cybersecurity. The more we rely on genome databases, the more likely these databases will become targets for cyber-attacks to interfere with public health and biosecurity systems by compromising their integrity, taking them hostage, or manipulating the data they contain. Also, while there is the potential to collect pathogen genomic data from infected individuals or agricultural and food products during disease outbreaks to improve disease modeling and forecast, how to protect the privacy of individuals, growers, and retailers is another major cyberbiosecurity challenge. As data become linkable to other data sources, individuals and groups become identifiable and potential malicious activities targeting those identified become feasible. Here, we define a number of potential cybersecurity weaknesses in today's pathogen genome databases to raise awareness, and we provide potential solutions to strengthen cyberbiosecurity during the development of the next generation of pathogen genome databases.

Keywords: cyberbiosecurity, cybersecurity, genome databases, pathogen, plant and animal health

1. INTRODUCTION

Current biological research, including pathogen related research projects, are increasingly dependent on public genome databases. Genome databases provide information about genomic sequences (Benson et al., 2018), gene annotations (Aken et al., 2016), protein sequences (Punta et al., 2012), protein interactions, and metabolic networks, which are playing crucial roles in designing and implementing biological experiments in many organisms.

A few key online databases provide repositories of raw data, processed data, and metadata generated by genome-scale sequencing projects (Leinonen et al., 2011a,b). Many specialized databases, such as pathogen-related databases (Winnenburg, 2006; Aurrecoechea et al., 2017; Wattam et al., 2017), provide curated data that serve specific research domains. In the 2018 Nucleic Acids Research (NAR) database issue, an online molecular biology database collection, 1,737 databases were reported as being publicly available. This article will review cybersecurity aspects of online genome databases with a focus on pathogen-related databases. Among the databases collected by NAR, 30 are dedicated to viral genomes, 71 to prokaryotic genomes, and 35 to fungal genomes. These databases are of great interest to pathogen research. Many general-purpose databases also contain data related to pathogen genomes, genes, and protein annotations (Mukherjee et al., 2017). Some online databases not only provide repositories of research data but also provide computational tools that allow users to perform genomic data analysis online (Wattam et al., 2017). As metagenome and transcriptome sequencing become common practice in pathogen research, online databases are important tools for annotating and interpreting these genome-scale experiments.

There has been an increasing number of high-profile cybersecurity breaches in recent years that have raised public awareness of potential social, political, and economic consequences that can be caused by such attacks (Newman, 2018). For example, private health record systems at hospitals have been targets for ransomware attacks in recent years (Osborne, 2018). However, cybersecurity awareness is still lacking in the research and health care industries (Kruse et al., 2017). Despite the importance of online genome databases to biological and pathogenic research, there is limited discussion, and virtually no research that focuses on biosecurity and cybersecurity risks (“cyberbiosecurity”) with regard to online biological databases. We suspect that one reason is that genome databases are most utilized by the research community. The number of people that can be directly affected by cyberattacks on genome databases is currently relatively small as compared to web sites or databases for large enterprises that have millions of users. Because of this perceived limited utilization of genomic data, there is limited incentive to target genome databases. However, given the millions of research dollars that are invested in generating genomic data yearly, it is surprising to see that there is almost no research that has been published related to protecting such data from cyberattacks.

Analogous to the tens of thousands of public libraries that hold the knowledge of humanity in the format of text books, public genome databases hold the entire body of genome research knowledge gained in the past thirty years. The size of public genomic data may someday surpass the size of all published text books combined. Besides the importance of protecting the products of public research investment, cyberbiosecurity research on genome databases is even more important because these databases contain so much of the knowledge gained over many years by the world-wide research community and because of the impact of this knowledge on human, animal, and plant health. Public genome databases also provide a unique resource for cyberbiosecurity research that aims to protect the bioeconomy (Murch et al., 2018; Peccoud et al., 2018), which has been estimated to consist in the USA of as much as 25% of GDP. The cyberinfrastructure and cybersecurity measures for the major biocompanies in health care, biopharma, and in the ag domains are largely unknown to the research community and thus cannot be easily analyzed. Unlike biocompanies, public genome databases have the intention to broaden their impacts by granting and facilitating open access to all users. Additionally, major innovations in computational methods for genomic data analysis are also largely driven by public research and open source software. Many companies, academic institutions, and government entities are likely also using open source software and databases developed in the public research community because the latest innovations in genomics are typically coming from academic research. Therefore, public genome databases provide a front-end of a highly innovative research community and are an ideal data resource for analyzing potential risks for cyberbiosecurity.

In the coming decades, we expect that genomic data will become more widely accessed and contextualized, and thus becoming increasingly relevant to public health and safety. For example, metagenomic sequencing can now be used to trace foodborne pathogen outbreaks (Huang et al., 2017; Kim et al., 2018), which potentially affect tens of thousands of consumers. Metagenomic sequencing is also used in detecting bacterial pathogens (Pendleton et al., 2017; Lazarevic et al., 2018), fungal pathogens (Tong et al., 2017), and viruses (Greninger et al., 2017; Lewandowska et al., 2017) in hospitals. Metagenomic sequencing is used in plant disease detection as well, such as detecting pathogens in wheat (Yiheng Hu, 2019) and other crops (Chalupowicz et al., 2019). For all these applications, reliable and accurate genome databases are essential for correct identification of disease-causing pathogens. A recent study has also shown that metagenomic sequencing can even reveal personal identity (Franzosa et al., 2015). Therefore, similar to the security and privacy concerns for personal genomic information (McGuire et al., 2008), personal metagenomic data is another area where cybersecurity is important in protecting sensitive, private, and health-related genomic data (Zmora et al., 2016). In this work, we present an overview of online pathogen genome databases (section 2), identify a number of potential cybersecurity weaknesses in today’s genome databases to raise awareness (section 3), and provide potential solutions to strengthen cybersecurity during the development of the next generation of genome databases (section 4). We focus on pathogen-related databases because of the direct health and agricultural implications of genomic studies of pathogens.

2. ONLINE DATABASES FOR PATHOGEN GENOME RESEARCH

Here, we provide an overview of existing databases that are related to pathogen genome research. We will explain the type of data that are hosted in these public genome databases, the potential usage of these data, and what will be the consequences if these databases are affected by cybersecurity breaches. We will review what types of access users can have to these databases and what the mechanisms are for users to contribute data. Finally, we will try to understand what cybersecurity measures will be needed to ensure the privacy, integrity, confidentiality, and availability of existing pathogen genome databases.

2.1. General Purpose Genome Databases With Pathogen Information

Most, if not all, molecular sequence data are deposited in the two major genomic data repositories: genome databases (Benson, 2004) hosted by the National Center for Biotechnology Information (NCBI) at the National Institutes of Health in the United States and genome databases (Hubbard et al., 2002) hosted at the European Molecular Biology Laboratory (EMBL). NCBI and EMBL provide databases for nucleotide sequences, protein sequences, genome assemblies, and genome annotations. Both databases also provide computational tools for users to query these databases through web-based interfaces or through programmatic access, such as using the command line or a programming language to access data stored in these databases. Here we focus on the resources and tools that are most relevant to pathogen genome research.

2.2. Sequence Databases at NCBI

The NCBI Assembly database (Kitts et al., 2016) is a database for assembled genomes of different organisms. This database hosts completed assemblies, contigs, scaffolds, and chromosomes. The database currently contains 4,055 fungal, 180,914 bacterial, and 23,816 viral genome assemblies (December, 2018). For each assembly, a summary page provides metadata and connections to other NCBI resources, such as its taxonomy browser (Federhen, 2012), original data in BioSample and BioProject (Barrett et al., 2012), and whole genome sequencing databases. Each assembly also includes detailed information regarding the identity of the data contributor. Both BioSample and BioProject are metadata repositories that save submitter-supplied metadata related to the nucleotide sequences and other data deposited at NCBI. More specifically, each BioSample typically includes descriptions of specific biomaterials, such as the name of a particular strain of bacteria. BioProjects are descriptions of larger projects, which consist of many BioSamples. Sequence data in NCBI are also organized under gene, EST, genome, nucleotide, and protein databases. RefSeq (O'Leary et al., 2016) is a database of annotated genes and genomes including many pathogens. Unlike other databases mentioned above, RefSeq provides genome annotations, so it is more useful for finding functional information for different pathogens. SRA (Leinonen et al., 2011b) is another NCBI database that hosts data from short read and long read sequencing projects. Data stored in the SRA database require extensive computational processing and analysis to convert them to useful whole genome sequences or transcriptome sequences that are more biologically meaningful. GEO data sets and GEO profiles (Clough and Barrett, 2016) are databases for gene expression and genome scale data related to gene regulations. These data were generated for various organisms using different technologies including RNA-seq, microarray, ChIP-seq, and other genomic experiments. GEO databases typically include both metadata and raw data from gene expression analysis. In summary, NCBI hosts dozens of databases that can be roughly characterized as the following: (1) sequence repositories, which include databases such as assembly, genome, gene, EST, nucleotide, and protein; (2) the RefSeq database that provides annotated sequences; (3) BioSample and BioProject databases that provide metadata for data sets deposited in the NCBI databases; (4) SRA that provides a repository of raw sequences requiring further processing to generate actual biologically meaningful data sets; and (5) the GEO database that provides genomic data sets related to regulation of gene expression.

Querying the NCBI databases is the most common usage in genomic pathogen studies. To search data from the NCBI database, a unified web interface is provided that allows querying all databases by any anonymous user. NCBI also allows a user to compose and use URLs to directly retrieve data from some databases. Many data sets in NCBI can also be downloaded anonymously using its FTP server. Programmable access is allowed through software developed by NCBI, including Entrez Programming Utilities (E-utilities) and the SRA-toolkit. E-utilities is a set of software tools that allows users to query the NCBI databases from a command line interface. A user is recommended to perform no more than 3 queries per second, otherwise the IP address of the user will be blocked. If the user's IP is blocked, the user must register through NCBI by providing additional information such as their email address and the tool name that the user is to develop. The reason for requiring a tool name is that some users (bioinformaticians) are interested in developing batch query tools for NCBI data. Starting in December 2018, an API (Application Programming Interface) key is required to perform more than 3 queries per second using E-utilities. An API key can be obtained by registered users of NCBI and is associated with a unique user account. Besides E-utilities, NCBI provides BLAST, which allows the user to query its sequence databases using nucleotide or protein sequences as inputs. SRA-toolkit is a utility software designed for downloading data from the NCBI-SRA database. This is because the data sets deposited in the SRA database are typically much larger than most other types of data sets in NCBI databases. No registration is required to use SRA-toolkit to download data, and there are no clear instructions on whether IP addresses will be blocked for using SRA toolkit if a certain bandwidth is exceeded. Queries sent to a database present a minimal risk to the stored data; however, vulnerabilities in the host system or database interpreter are subject to exploitation when vulnerabilities are discovered in the query interpreter. Typically, systems are targeted for attack in order to escalate operating privileges on the host itself, so that its resources can be redirected for the attackers purpose to become a platform for attacks on other systems, or to monetize computing resources by generating digital coinage.

Submitting data to the NCBI database is a multi-step, well-controlled process. For example, to submit a genome assembly to the Assembly database, detailed instructions are provided on the NCBI web site. First, the user is required to login to the BioProject portal using preregistered credentials to establish a project ID and fill out a submission template file. After obtaining the project ID, the user needs to organize the data and metadata of the project into specific formats, such as FASTA and AGP formats. Some data files have to be converted to a specific format using software utilities developed by NCBI.

Once the data files are in the correct format, the user will use another web portal from NCBI to submit and upload the data. After submission, the user will have to send an email to an administrator account at NCBI that includes the description of the project. The process of submitting data to the SRA database is similar to the above described procedures. Because the files that are uploaded to SRA are typically much larger than other data types, the user can use FTP and Aspera Connect to upload files to a predefined FTP folder provided by NCBI administrators by email.

2.3. Sequence Databases at EMBL

Genome databases managed by EMBL are mainly through the European Bioinformatics Institute (EBI). Similar to GeneBank (Benson et al., 2018) at NCBI, the European Nucleotide Archive (ENA) (Leinonen et al., 2011a) is a repository of sequence information, including those of pathogens. Some data types are the same between ENA and NCBI, such as assembly and EST data sets. However, these data sets have different sequence identifiers in EMBL compared to NCBI databases. Some data types are similar between ENA and NCBI. For example, ENA uses Sample and Study in place of BioSample and BioProject in NCBI. Some data types are unique to ENA, such as CDS data sets that are found in ENA only. Another major genome database from EMBL is the EnsemblGenomes database (Hubbard et al., 2002). There are multiple databases such as Ensembl Bacteria and Ensembl Fungi that are most relevant to pathogen research. Ensembl Bacteria includes genomes of 44,048 bacterial species, and Ensembl Fungi includes genomes of 811 fungal species. Both databases provide gene and genome annotations. Transcriptome data are available from multiple databases in EBI. ArrayExpress (Kolesnikov et al., 2015) is a database that contains gene expression data and results from other functional genomic assays. Although the name ArrayExpress suggests that the database contains data generated by microarray analysis, the database actually contains RNA-seq, DNA-seq, ChIP-seq, and methylation data, which makes this database very similar to the GEO database in NCBI. Expression Atlas (Papatheodorou et al., 2018) is a curated database for gene expression data only. In regard to raw data of sequencing experiments, SRA also has its counterpart in EMBL, which is also called SRA but is part of the ENA database.

To query data from EMBL databases, a number of methods are available. All databases (ENA, Ensembl genomes, ArrayExpress and Expression Atlas) support text-based queries. ENA also allows sequence-based searches. Several databases provide programmatic access through the REST interface, which allows the user to retrieve data using a URL following a specific syntax. A user can also perform sequence-based searches using REST and SOAP APIs (Application Programming Interface, a commonly shared set of procedures for accessing data across different software platforms). There is a limit of 30 queries at a time if a user uses REST or SOAP to access data. ArrayExpress, ENA, Ensembl Genome, and Expression Atlas all provide FTP access to users for bulk download purposes. Some databases provide additional options for data download. For example, Ensembl Genome databases allow the user to download data by downloading a MySQL dump through their FTP site. Ensembl genome databases also allow users to directly access the MySQL database server with a MySQL client, or by using a PERL API to access a MySQL database. Finally, Biomart (Smedley et al., 2009) is another interface for data access to EMBL databases. A user can use a web interface to interact with Biomart to retrieve data from EMBL databases. Alternatively, a user can use REST, MySQL, a PERL API, or an R API to access data from Biomart.

To submit data to any of the EMBL sequence databases requires processes that are similar to submitting data to NCBI. For example, if a user wants to submit a data set to the ArrayExpress database, the user has to first register an account associated with a user-provided email address and password. The user has to prepare metadata, raw data, and processed data according to a specific format, as required by ArrayExpress. A web interface called Annotare provides detailed, step-by-step instructions on how to upload data through the Annotare web interface.

In this section, we reviewed two of the largest molecular databases in the world, found at NCBI and EMBL. There are several properties that these two databases have in common. Both databases host terabytes of genomic data in many specific data formats. Types of data include metadata, raw data, and processed data. Some data are in text files with specific structures. For example, biological sequences are stored in FASTA and FASTQ format, which are specifically designed for storing molecular sequences, and the correctness of these formats can be checked automatically by computer programs. Some data are binary data, such as SRA data, which require software tools to extract information into human-readable formats. Users can access data using a multitude of methods, including web interface, PERL or R API, MySQL query, REST URL, SOAP, and FTP download. Certain download limits are implemented in both databases to limit the amount of data or the speed of data download by users. To submit data to these databases, a user will need to preregister with an email address and login to use a site-specific web interface to upload data. Metadata, raw data, and processed data can be uploaded, and web forms will need to be filled out to describe the data. Although not explicitly stated in the guideline of the submission process, for both web sites, there are curators that control the final process of integrating user-submitted data into the database.

2.4. JGI Genome Databases

The integrated genome and metagenome comparative data analysis system (IMG/M) (Chen et al., 2017) is a database containing tools to annotate microbial genomes and metagenomes. MycoCosm (Grigoriev et al., 2014) is a web portal that hosts fungal genome data. Genomes OnLine Database (GOLD) (Mukherjee et al., 2017) is a database that manages metadata and raw data for genome and metagenome sequencing projects. IMG/M, MycoCosm, and GOLD are all developed by the Joint Genome Institute (JGI) and are supported by the Department of Energy. A large fraction of data in these databases are imported from NCBI Gene Bank and other related databases discussed above. The three JGI databases are unique, because a substantial portion of their genome data are generated by JGI itself. These databases also provide a repository for metagenome sequencing projects, computational tools for gene and genome annotation, and comparative genome analysis. These features are important for functional analysis of microbial genomes but are not clearly present in the NCBI databases or EMBL-EBI genome databases.

2.5. Other Specialized Microbial and Pathogen Databases

In addition to the major sequence repositories described above, there are many databases and web services that are of smaller scale and have more specific focus on certain aspects of pathogen genomics. We introduce some of these databases as examples to discuss potential cybersecurity concerns of these databases.

The Pathosystems Resource Integration Center (PATRIC) is a bacterial bioinformatics center (Wattam et al., 2017) that was first established by the National Institute of Allergy and Infectious Diseases (NIAID) as the National Microbial Pathogen Data Resource (NMPDR) (NMPDR, 2019). The major focus of PATRIC is bacterial genome annotation and analysis. There are 202,602 bacterial genomes hosted at PATRIC currently; however, there are also thousands of archaea and phage genomes available in PATRIC. PATRIC provides users with resources in genome, transcriptome, protein interaction, protein structure, signaling pathways, and metabolic pathways annotations for these genomes. Metadata annotation for these data resources are also available. PATRIC also provides analytic tools that allow any user to perform genome assembly, genome annotation, proteome comparisons, RNA-seq analysis, variant analysis, and metabolic pathway model reconstruction. PATRIC also allows users to upload their own gene expression data to perform analysis online. To use these services at PATRIC, a user needs to provide an email address, which will receive a link to a page that allows users to set up a password. Once logged in, the user can use a web-based interface to define their analytic pipelines using a number of published software packages.

PATRIC represents a very common model of genome databases. First, the PATRIC database collects data from multiple external resources including NCBI GeneBank, genome sequencing centers, and other collaborators. Second, a unified pipeline was developed to provide annotation to the sequence data and the processed results are stored in the PATRIC database. Metadata is curated by the PATRIC team and also deposited in the PATRIC database. Third, an external user can use computational tools and computing power provided by PATRIC to analyze user-generated data. To incorporate new user data into the PATRIC database, the user has to contact the PATRIC team through email, and the data will be curated by the PATRIC team before integration into the PATRIC database, although there are no defined industry standards for the curation activities.

There are many additional pathogen databases that are available, and here we provide a brief survey. The Eukaryotic Pathogen Genomics Database resource (EuPathDB) is a collection of databases for Eukaryotic pathogens, their related, non-pathogenic species, and selected host genomes (Aurrecoechea et al., 2017). EuPathDB provides genome, gene, protein, and metabolic pathway annotation as well as many other resources. EuPathDB also provides curated phenotypes, copy number variation, and polysomal transcriptomic data. EuPathDB allows users to build their analysis pipeline through a Galaxy workspace and some user-defined pipeline that can be made public. ViPR is a virus pathogen database (Pickett et al., 2012), which provides a web interface to search genome sequences, gene sequences, protein sequences, and protein structures. Online tools are provided for phylogenetic analysis, comparative genomic analysis, and genome annotation. PHI-base (Winnenburg, 2006; Urban et al., 2017) is a curated database for genes related to host-pathogen interactions. Currently, PHI-base contains information regarding 6438 genes and 11340 interactions between 263 pathogens and 194 hosts. PHI-base includes pathogen information for animal, plant, and fungal pathogens. A user must register before downloading data from PHI-base; however, searching the database does not require registration. PHIDIAS (Xiang et al., 2007) is a curated online database focused on genome, protein domain, and gene expression data related to pathogen and host interactions. Victors (Sayers et al., 2018) is a newly published system under the PHIDIAS database. The focus of Victors is on virulence factors and, currently, there are 5,296 virulence factors stored in the Victors database.

For plant pathogen related resources, PAMDB is a database and website for Plant-Associated Microbes (Almeida et al., 2010) and is designed to store and search data for multi-locus sequence typing for plant pathogenic bacteria. PhytoPath (Pedro et al., 2016) is an online database for genome data of plant pathogens. PhytoPath integrates a genome browser from Ensembl genomes and also provides links to PHI-base.

GenomeTrakr (Allard, 2016) is an FDA-led network of open source, whole genome sequencing projects that involves state, federal, international, and commercial partners. The goal of the GenomeTrakr project is to track food-borne pathogens through whole genome sequencing. One unique feature of the GenomeTrakr project is that there is no centralized data repository for this project hosted by FDA. Data generated from the GenomeTrakr project are deposited under the NCBI BioProject and SRA databases. Database for Reference Grade Microbial Sequences (FDA-ARGOS) is another FDA-led project that has generated high quality, reference-grade genomes for 2000 biothreat microorganisms and common clinical pathogens. The results of this project are also deposited as BioProjects in an NCBI database. These genome databases are summarized in **Table 1**.

As can be seen in this summary table, all databases described in this review provide metadata (**Table 1**, E) associated with sequence data (**Table 1**, A) and sequence data annotation (**Table 1**, B, D). Inclusion of standardized metadata in genome databases to facilitate data interpretation and data reuse has been

TABLE 1 | Functions of genomic databases.

Functions of online databases	Names of online databases
A. Contain genome, transcriptome, proteome sequences.	NCBI, EBI, DDBJ, JGI, PATRIC, EuPathDB, PAMDB, PHI-base, PHIDIAS, ViPR
B. Contain genome, transcriptome, proteome annotation.	NCBI, EBI, DDBJ, JGI, PATRIC, EuPathDB, PAMDB, PHI-base, PHIDIAS, ViPR
C. Provide raw data repository.	NCBI, EBI, DDBJ, JGI
D. Provide processed data.	NCBI, EBI, DDBJ, JGI, PATRIC, EuPathDB, PAMDB, PHI-base, PHIDIAS, ViPR
E. Include metadata.	NCBI, EBI, DDBJ, JGI, PATRIC, EuPathDB, PAMDB, PHI-base, PHIDIAS, ViPR
F. Include single purpose bioinformatics tools such as BLAST as a service or query tool.	NCBI, EBI, DDBJ, JGI, PATRIC, EuPathDB, PAMDB, PHI-base, PHIDIAS, ViPR
G. Include analysis pipeline build.	PATRIC, EuPathDB, ViPR
H. Upload data access control.	NCBI, EBI, DDBJ, JGI, PATRIC, EuPathDB, PAMDB, PHI-base, ViPR
I. Complete download data access control.	JGI, PAMDB
J. Require strong password.	None
K. Allow programmatic access.	NCBI, EBI, DDBJ, JGI (Globus), PATRIC, EuPathDB

a major focus of the genomic research community in the last two decades (Brazma et al., 2001; Brazma, 2009). Because of the awareness of the importance of metadata, including metadata has become a standard for current genome databases. We also found that only four major databases contain raw data repositories and some raw data can only be found in a single raw data repository. This is because maintaining large amounts of raw sequence data is cost-prohibitive for smaller institutions. However, the current situation does introduce a high risk of data loss in the event that one of these raw data repositories is disrupted and redundancy measures fail, resulting in substantial data loss.

Most databases allow the use of bioinformatics tools such as BLAST, which is used to perform similarity-based queries of a genome database with user-provided input sequences (**Table 1**, F). There are a limited number of such tools that are available and most of these tools are open source and have been widely used. Although security issues with these tools have not been reported, even if such security risk were to exist, it is relatively easy to control by dedicated measures. For example, funding could be provided to qualified individuals or entities to routinely check the security risk of these few, widely used computational tools. However, what concerns us more is that there is a growing number of a new generation of databases, which provide the users with the capacity to build customized analytical pipelines, composed of distributed compute and storage resources across multiple physical and virtual systems of unknown integrity. Unlike BLAST, some computational tools used in these customized pipelines may not be widely scrutinized. As the genomic data analytics community continuous to grow, more highly specialized new tools are likely to emerge. Data processing pipelines composed of many newly developed computational tools are more susceptible to contain intentional or unintentional vulnerable code or shared libraries and may be much more difficult to maintain and may become more difficult to mitigate security risks.

Another important feature of these databases is that all databases require access control when users request data upload to the main database (**Table 1**, H). There is always "a human in the loop" to curate and authenticate the user before data are integrated into the database, although there also could still be risk associated with large data uploads when a complete sanity check is computationally prohibitive (see next section). In contrast to upload control, one concern is that only two databases ask for complete data access control (**Table 1**, I). Complete access control means that a user must register and then login before downloading any data from a database. Most databases provide anonymous download without any control, while some databases (such as NCBI and EBI) do provide throttle mechanisms to curb rapid download of multiple records. Finally, the most concerning problem is that none of the databases reviewed here requires strong passwords, which may lead to multiple cybersecurity risks (see next section).

Finally, some of the databases provide methods for programmatic access, which is to help the users to perform structured queries with programming languages (PERL API) or relational databases (SQL query), or provide faster download speed with external services or fast downloading protocols such as globus and ascp. The risk of using these third party software tools is related to each individual software and can be mitigated accordingly. Since many of these tools are broadly used outside the genomic research community, a simple way to mitigate risk is to raise awareness of genomic research programmers and database managers in the security risk announcements for these computational tools.

3. SECURITY THREATS

Cybersecurity broadly focuses on the confidentiality, integrity, and availability of digital information (Jang-Jaccard and Nepal, 2014) of all types, including genomic data. Yet there has not been a systematic study concerning security breaches of genome databases. However, personal medical information subjected to ransomware attacks has been reported (Kruse et al., 2017). This topic is not within the scope of this review. Although there is no public report for security breaches of molecular databases, existing cyberattack methods could easily target current molecular databases. We discuss the potential damages that can be caused by cyberattacks to genome databases as summarized in **Table 2**.

Confidentiality. One major motivation behind cyberattacks is to gain access to sensitive personal information. Most public genome databases do not contain sensitive personal information such as credit card numbers or social security numbers, yet they do contain individual's genomic data, perhaps the most

TABLE 2 | General security threats for genome databases.

Threat	Impact	Remedy
Confidentiality	Privacy of individuals, leaking credentials	Encryption, strong authentication, access control, data anonymization
Data Integrity	Invalid data	Strong identity verification (such as the use of certificates), encryption, checksum verification
Data Availability	Query performance, denial of service	Distributed data providers, intrusion detection and prevention

"personal" data of all. Two reasons that genome databases have not been targeted by cyberattacks is that (1) the population of users of genome databases are mainly research scientists, and accounts for a small percentage of the entire population; and (2) the technology needed to exploit the data has been sophisticated and expensive. However, growth in the field has led to both of the factors eroding in impact. As knowledge and training spreads, and with technological advances, the equipment becomes less expensive and easier to use. Additionally, indiscriminate attacks can always happen and can cause damage to genome databases. A common vulnerability found in this review is that while many databases do require user email and password to establish access control, users repeat their email and password combinations. Thus, credentials compromised in one system could be of interest to attackers to gain access to other accounts of the same user elsewhere. Among the databases we have reviewed, almost no database requires strong passwords, i.e., mandating a sufficiently long password of sufficient complexity (that includes capital letters, numbers, and symbols) to make brute force account password attacks impractical.

A general approach to data confidentiality is to secure the database using methods to maintain data privacy (Bajaj and Sion, 2014). A method with a growing interest is to use encrypted databases (Ravan et al., 2013) with proper access control and high assurance encryption standard. Protecting against privacy attacks, existing methods such as *k*-anonymity (Samarati and Sweeney, 1998; Zhong et al., 2005) can be utilized. Data anonymization can be a challenging task and depends on the structure of the data. Note that methods for data de-anonymization have been suggested (Narayanan and Shmatikov, 2008).

Another major issue with genome databases is using the idea of correlation attacks (Meier and Staffelbach, 1989). The attacker wishes to correlate biological data to specific users or groups of users. The threat can be from authenticated and/or unauthenticated malicious clients. In the first case, an authenticated client is one that has access to the database and can read and correlate records in multiple databases. This is typically referred to as an insider threat and requires a vigilant user review and monitoring process to identify potential candidates. In the second case, the attacker uses a classical external attack, for example exploiting an existing user's credentials, or sending emails to known system users with malware embedded (also known as "phishing") to gain access to system accounts and then proceed with a correlation attack.

One unique concern for pathogen genome databases is that the knowledge of pathogen sequences may lead to malicious use. Such ill-intended use of genomic data and technology is a major biosecurity concern. Currently, many genomes of animal and plant pathogens are freely accessible to any user through pathogen genome databases. A study in early 2000s had concluded that open access to pathogen genomes should be promoted (Committee on Genomics Databases for Bioterrorism Threat Agents et al., 2004). However, situations have changed due to the reduced cost of synthetic DNA technology and advancement in synthetic biology (Hughes and Ellington, 2017). Even the genome sequence of such a high-risk pathogen as the smallpox virus, Variola major, can be easily accessed at NCBI by any anonymous user. Putting in place more stringent regulations regarding access to sensitive data by governments is one potential solution for this problem. However, it is challenging to determine what should be regulated and what should not be regulated, particularly in a collaborative research setting. Imposing such regulations may also discourage research groups to conduct research related to these pathogens and increase the operating costs for those groups. In our opinion, one possible model is that, instead of granting free access to pathogen genomes to anyone who has an internet connection, funding agencies could control the access to genomic information for high-risk pathogens. Genomic data for these pathogens would only be available once the corresponding grant application has been peer-reviewed and determined as fund-able.

Data integrity. Genome databases grow rapidly due to the increasing amount of sequencing data. Many genome databases have protocols for data quality control and manual curation, which are two methods to ensure data integrity. For all databases reviewed in this article, to submit a new data set to these databases, a user has to register an account with an email address and the data submitted to the database cannot be directly inserted in the main database. There is always a curator or an administrator to oversee the process. In several cases, a user can upload his own data to the server and perform analysis using the web interface provided by the database. However, user-provided data cannot be directly integrated into the main database in any situation. Many web sites provide methods for users to upload data. Interestingly, there seems to be no case where the data integrity is checked during the transfer process to ensure that the data provided by the user is not modified during the data transfer process. The rapid growth of the genomic and bioinformatic fields has also created a volume processing challenge for curators, where data science has introduced database sizes in the peta- and exa-byte range that has left institutions scrambling to bring massive "big data" computing infrastructures on-line and growing at a schedule that keeps pace with the growth of available data. Almost all traditional cybersecurity solutions fail at data volume, velocity, and variety of this scale.

Attackers have several options to exploit an unverified data transfer process. One possibility of attack is to provide invalid data, motivated to guide future studies toward specific outcomes. This attack requires careful crafting of records in the database to maintain a valid format but containing data without experimental evidence. This attack can be done during a single data transfer. The mitigation is to use thorough analysis of the data at transfer time. The analysis can easily ensure correct formatting of the data, discarding garbled input. However, verifying the validity of the data is particularly challenging and cannot be easily performed using existing methods. Another type of attack consists in gradually injecting invalid records within a larger valid data set. For example, the attacker could download existing data from the database, extract a subset of the data, and inject invalid input. In this case, detection mechanisms that use probabilistic analysis can fail to find the invalid records. Only records with clear violation of data integrity can be detected. Such attacks have been proposed previously in various contexts, for example (Mo et al., 2010; Cárdenas et al., 2011; Esmalifalak et al., 2013).

Data availability. Reduced data availability is a potential concern for genomic databases. This will cause delay in progress for time-sensitive experiments. For example, if a diagnostic lab is using DNA sequences as a method to identify pathogens, disruption of a database will cause delays in obtaining an identification. However, it is hard to estimate how many research projects or clinical operations do require real-time query of remote servers or databases. One major reason for loss of data availability is that web sites or databases are no longer maintained. In some cases, an older version of a website (NMPDR) is superseded by a new website (PATRIC). In several cases (not listed in this article), the web site link simply becomes obsolete. Another reason for loss of data is the adoption of distributed data models across shared high-performance research networks. A database may be freely shared in such a collaboration space, but there are not always resources to keep the database online and available for sharing with future users. It is hard to estimate the impact of such loss of availability. However, the manual labor for data curation, system and data integration, and web site development are lost.

To maintain data availability, a distributed network of permanent data providers is needed (Jsang et al., 2007). The network can include centralized control systems that can provide freshness guarantees and can maintain availability when some data providers are no longer responsive. The associated monetary cost, performance issues, and organizational aspects of this network require careful considerations.

Attack on physical hardware. In some databases that we have reviewed, MySQL query, a REST API, and a PERL API are provided for remote users to query data directly. In these scenarios, the databases are susceptible to attacks such as SQL injection. However, there is a limited public record of how many genome databases have been attacked by these means. Several databases provide computational tools to annotate microbial genomes, perform genome assembly, and search genome database. These computational analyses typically require substantial computing power. Many major research universities are equipped with cluster computing servers that have been used as the backends for these computationally intensive services. Therefore, these servers are attractive targets for malicious usage such as mining of cryptocurrency (Tahir et al., 2017).

Future physical exploitation. As genomic data become an integral part of an individual's healthcare and treatment plan, the traditional firewall between bioinformatics and medical technology becomes more porous. Thus, an acceptable operating risk for a genome database may be transmitted downstream, where it becomes an unacceptable threat to the technology responsible for a patient's care. As this threat scenario evolves from the hypothetical to the possible, today's low risk research data will become the foundation for the high risk and critical health care analyses; security controls that today seem to lack a Return On Investment (ROI) for their overhead costs will have to be retrofitted, or the whole body of work will have to be revalidated and secured properly.

In summary, genome databases will only grow as a target for multiple existing cybersecurity threats and threat actors. Users of genome databases could lose their personal information, such as an email address and associated password. Patients and subjects of research that capture genomic data may find that institutions have lost control of the most intimate data available about them. Many genome databases have established local standards for data quality and metadata curation and include administrators to oversee the process of data upload but have not engaged with current cybersecurity best practices and industry standards to protect the systems and networks upon which they rely. Data availability can affect users productivity but the actual costs of malicious cyber activity in the genomics field is difficult to quantify, and no one has accurately developed a methodology for financial loss estimates. Yet the lack of a measure of the risk does not negate the very real risks that exist.

4. SECURITY REQUIREMENTS AND POTENTIAL FOR NEW APPROACHES

In this section, we discuss the existing security measures used by pathogen genome databases and what can be potentially improved in current practices.

Access control. Many databases reviewed in this article contain components that do not require login. Users can simply use a web interface to query data from databases such as NCBI, EMBL-EBI, and many other specialized databases. Users also can use a programming language, a REST API, or a MySQL query to access data. For batch download, anonymous FTP access is provided in several cases. Both, NCBI and EMBL-EBI, have implemented speed limits for bulk download using programmable interfaces. IP block is used by NCBI to limit download speed. Almost all databases require users to use email and password as methods for login to gain access to data upload and data analytic capability. We notice that most databases do not require strong passwords, such as combinations of long phrases, capital letters, symbols, and number. No databases reviewed in this article require two-factor authentication or login through third party accounts. Requiring strong passwords, implementing two-factor authentication, and implementing login through third party accounts (Google, ORCID, or institution-specific accounts) could provide additional security measures for the current generation of genomic databases.

Database access control systems is a well studied subject (Bertino et al., 1996; Kalam et al., 2003), with implications for mobile (Xu et al., 2016) and web applications (Xu et al., 2017). Classical access control systems, for example access control matrix, could be used (Sandhu, 1992) to provide basic functionality for systems that interface genomic databases. The core challenge here is to accommodate special use cases that are standard in the genomic research community, maintaining usability and performance while achieving high assurances.

Data integrity check and protection. Most databases allow users to contribute data and implement metadata standards. However, it is unclear how databases ensure that the data are intact during the transfer process. Simple methods such as cryptographic checksums could be implemented to ensure data integrity. There are concerns that malicious users can inject large amounts of useless data to public databases. Current quality control mechanisms do not allow the curators to control data quality with regard to the above mentioned, hypothetical situation. However, a large data set upload is controlled by the database administrator such that it is unlikely that random, large data sets can be integrated into a database without being noticed. Another possibility is that malicious users can modify certain records in a public database. However, it is difficult to imagine the motivation for performing such an attack on public genome databases. Another model for data protection is the use of encrypted databases (Eykholt et al., 2017) or the use of secure multiparty computation (Evans et al., 2018). For example, access to databases does not have to be binary, allowing or denying access based on access control models. One can reveal partial views of the database as needed.

Data availability and longevity. Loss of database access entails loss of valuable research results and waste of manual labor in the data curation process. Since maintenance of online databases requires continuous manual support, it is common that some databases cannot be maintained due to lack of funding support. One solution to this problem is to deposit data to public databases that are maintained by national governments such as the databases managed by NCBI. Two examples we reviewed are FDA-ARGOS and FDA-GenomeTrakr projects. Neither of these projects maintain their own databases for the data generated by these projects. Instead, data are uploaded as BioProjects to the NCBI database. This approach provides better guarantee for longer term availability of research data. NCBI BioProject and GEO databases provide a good repository for genomic data and these databases are not limited to the deposit of raw data alone.

5. CONCLUSIONS

In the past 30 years, pathogen genome databases and genome databases in general have become an integral part of biological and biomedical research. Although genome databases have not been reported as primary targets of cybersecurity threats, many common cybersecurity threats are applicable to genome databases. Disrupting genome databases can lead to loss of productivity, loss of research investment, and loss of private data, such as email addresses and passwords. Computing servers used by genome databases can be hijacked for cryptocurrency mining or other malicious purpose. Since the revolution of genomic science started by sequencing human genomes, billions of research funding have been invested in performing genomic experiments, generating genomic data, annotating, curating, and interpreting genomic data. Despite this large investment in genomic sciences, we found there is almost no dedicated research that focuses on protecting such data from cybersecurity threats. We think that it is necessary for the community that develops genomic databases to collectively design a minimum, necessary security standard for new genome database projects.

AUTHOR CONTRIBUTIONS

SL developed the first draft of the manuscript. BV, LH, HA, MS, and CL provided comments and made edits to the manuscript.

FUNDING

This research was supported by seed funds for cyberbiosecurity from the College of Agriculture and Life Sciences at Virginia Tech. SL and BV was also provided in part by the Virginia Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture, US Department of Agriculture.

ACKNOWLEDGMENTS

The findings and conclusions in this preliminary publication have not been formally disseminated by the U.S. Department of Agriculture and should not be construed to represent any agency determination or policy.

REFERENCES

Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., et al. (2016). The Ensembl gene annotation system. *Database* 2016:baw093. doi: 10.1093/database/baw093

Allard, M. W. (2016). The future of whole-genome sequencing for public health and the clinic. *J. Clin. Microbiol.* 54, 1946–1948. doi: 10.1128/JCM.01082-16

Almeida, N. F., Yan, S., Cai, R., Clarke, C. R., Morris, C. E., Schaad, N. W., et al. (2010). PAMDB, A Multilocus Sequence Typing and Analysis Database and Website for plant-associated microbes. *Phytopathology* 100, 208–215. doi: 10.1094/PHYTO-100-3-0208

Aurrecoechea, C., Barreto, A., Basenko, E. Y., Brestelli, J., Brunk, B. P., Cade, S., et al. (2017). Eupathdb: the eukaryotic pathogen genomics database resource. *Nucleic Acids Res.* 45, D581–D591. doi: 10.1093/nar/gkw1105

Bajaj, S., and Sion, R. (2014). Trusteddb: a trusted hardware-based database with privacy and data confidentiality. *IEEE Trans. Knowl. Data Eng.* 26, 752–765. doi: 10.1109/TKDE.2013.38

Barrett, T., Clark, K., Gevorgyan, R., Gorelenkov, V., Gribov, E., Karsch-Mizrachi, I., et al. (2012). BioProject and BioSample databases at NCBI: facilitating capture and organization of metadata. *Nucleic Acids Res.* 40, D57–D63. doi: 10.1093/nar/gkr1163

Benson, D. A. (2004). GenBank. *Nucleic Acids Research*, 33, D34–D38. doi: 10.1093/nar/gki063

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., et al. (2018). GenBank. *Nucleic Acids Res.* 46, D41–D47. doi: 10.1093/nar/gkx1094

Bertino, E., Jajodia, S., and Samarati, P. (1996). "Supporting multiple access control policies in database systems," in *Proceedings 1996 IEEE Symposium on Security and Privacy* (Los Alamitos, CA), 94–107.

Brazma, A. (2009). Minimum information about a microarray experiment (MIAME) successes, failures, challenges. *Sci. World J.* 9, 420–423. doi: 10.1109/SECPRI.1996.502673

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., et al. (2001). Minimum information about a microarray experiment (MIAME) toward standards for microarray data. *Nat. Genet.* 29, 365–371. doi: 10.1038/ng1201-365

Cárdenas, A. A., Amin, S., Lin, Z.-S., Huang, Y.-L., Huang, C.-Y., and Sastry, S. (2011). "Attacks against process control systems: Risk assessment, detection, and response," in *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '11 (New York, NY: ACM), 355–366.

Chalupowicz, L., Dombrovsky, A., Gaba, V., Luria, N., Reuven, M., Beerman, A., et al. (2019). Diagnosis of plant diseases using the nanopore sequencing platform. *Plant Pathol.* 68, 229–238. doi: 10.1111/ppa.12957

Chen, I.-M. A., Markowitz, V. M., Chu, K., Palaniappan, K., Szeto, E., Pillay, M., et al. (2017). IMG/M: integrated genome and metagenome comparative data analysis system. *Nucleic Acids Res.* 45, D507–D516. doi: 10.1093/nar/gkw929

Clough E., and Barrett, T. (2016). "The gene expression omnibus database," in *Statistical Genomics. Methods in Molecular Biology,* Vol. 1418, eds E. Mathé and S. Davis (New York, NY: Humana Press).

Committee on Genomics Databases for Bioterrorism Threat Agents, Board on Life Sciences, D. o. E., Life Studies, P., and Division, G. A. (2004). *Seeking Security*. Washington, DC: National Academies Press.

Esmalifalak, M., Nguyen, N. T., Zheng, R., and Han, Z. (2013). "Detecting stealthy false data injection using machine learning in smart grid," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 808–813. doi: 10.1109/GLOCOM.2013.6831172

Evans, D., Kolesnikov, V., and Rosulek, M. (2018). A pragmatic introduction to secure multi-party computation. *Found. Trends Privacy Secur.* 2, 70–246. doi: 10.1561/3300000019

Eykholt, K., Prakash, A., and Mozafari, B. (2017). "Ensuring authorized updates in multi-user database-backed applications," in *26th USENIX Security Symposium (USENIX Security 17)*, eds E. Kirda and T. Ristenpart (Vancouver, BC: USENIX Association), 1445–1462.

Federhen, S. (2012). The NCBI Taxonomy database. *Nucleic Acids Res.* 40, D136–D143. doi: 10.1093/nar/gkr1178

Franzosa, E. A., Huang, K., Meadow, J. F., Gevers, D., Lemon, K. P., Bohannan, B. J. M., et al. (2015). Identifying personal microbiomes using metagenomic codes. *Proc. Natl. Acad. Sci. U S A.* 112, E2930–E2938. doi: 10.1073/pnas.1423854112

Greninger, A. L., Zerr, D. M., Qin, X., Adler, A. L., Sampoleo, R., Kuypers, J. M., et al. (2017). Rapid metagenomic next-generation sequencing during an investigation of hospital-acquired human parainfluenza virus 3 infections. *J. Clin. Microbiol.* 55, 177–182. doi: 10.1128/JCM.01881-16

Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., et al. (2014). MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic Acids Res.* 42, D699–D704. doi: 10.1093/nar/gkt1183

Hu, Y., Green, G. S., Milgate, A. W., Stone, E. A., Rathjen, J. P., and Schwessinger, B. (2019). Pathogen detection and microbiome analysis of infected wheat using a portable DNA sequencer. *bioRxiv*. [Preprint]. doi: 10.1094/PBIOMES-01-19-0004-R

Huang, A. D., Luo, C., Pena-Gonzalez, A., Weigand, M. R., Tarr, C. L., and Konstantinidis, K. T. (2017). Metagenomics of two severe foodborne outbreaks provides diagnostic signatures and signs of coinfection not attainable by traditional methods. *Appl. Environ. Microbiol.* 83, e02577–16. doi: 10.1128/AEM.02577-16

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., et al. (2002). The Ensembl genome database project. *Nucleic Acids Res.* 30, 38–41. doi: 10.1093/nar/30.1.38

Hughes, R. A., and Ellington, A. D. (2017). Synthetic DNA synthesis and assembly: putting the synthetic in synthetic biology. *Cold Spring Harbor Perspect. Biol.* 9:a023812. doi: 10.1007/978-1-4939-6343-0

Jang-Jaccard, J., and Nepal, S. (2014). A survey of emerging threats in cybersecurity. *J. Comput. Syst. Sci.* 80, 973–993. doi: 10.1016/j.jcss.2014.02.005

Jsang, A., Ismail, R., and Boyd, C. (2007). A survey of trust and reputation systems for online service provision. *Decision Supp. Syst.* 43, 618 – 644. doi: 10.1016/j.dss.2005.05.019

Kalam, A. A. E., Baida, R. E., Balbiani, P., Benferhat, S., Cuppens, F., Deswarte, Y., et al. (2003). "Organization based access control," in *Proceedings POLICY 2003. IEEE 4th International Workshop on Policies for Distributed Systems and Networks* (Los Alamitos, CA), 120–131. doi: 10.1109/POLICY.2003.1206966

Kim, D., Hong, S., Kim, Y.-T., Ryu, S., Kim, H. B., and Lee, J.-H. (2018). Metagenomic approach to identifying foodborne pathogens on chinese cabbage. *J. Microbiol. Biotechnol.* 28, 227–235. doi: 10.4014/jmb.1710.10021

Kitts, P. A., Church, D. M., Thibaud-Nissen, F., Choi, J., Hem, V., Sapojnikov, V., et al. (2016). Assembly: a resource for assembled genomes at NCBI. *Nucleic Acids Res.* 44, D73–D80. doi: 10.1093/nar/gkv1226

Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., et al. (2015). ArrayExpress update-simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116. doi: 10.1093/nar/gku1057

Kruse, C. S., Frederick, B., Jacobson, T., and Monticone, D. K. (2017). Cybersecurity in healthcare: a systematic review of modern threats and trends. *Technol. Health Care* 25, 1–10. doi: 10.3233/THC-161263

Lazarevic, V., Gaïa, N., Girard, M., Leo, S., Cherkaoui, A., Renzi, G., et al. (2018). When bacterial culture fails, metagenomics can help: a case of chronic hepatic brucelloma assessed by next-generation sequencing. *Front. Microbiol.* 9:1566. doi: 10.3389/fmicb.2018.01566

Leinonen, R., Akhtar, R., Birney, E., Bower, L., Cerdeno-Tarraga, A., Cheng, Y., et al. (2011a). The european nucleotide archive. *Nucleic Acids Res.* 39, D28–D31. doi: 10.1093/nar/gkq967

Leinonen, R., Sugawara, H., and Shumway, M. (2011b). The sequence read archive. *Nucleic Acids Res.* 39(Suppl. 1), D19–21. doi: 10.1093/nar/gkq1019

Lewandowska, D. W., Schreiber, P. W., Schuurmans, M. M., Ruehe, B., Zagordi, O., Bayard, C., et al. (2017). Metagenomic sequencing complements routine diagnostics in identifying viral pathogens in lung transplant recipients with unknown etiology of respiratory infection. *PLoS ONE* 12:e0177340. doi: 10.1371/journal.pone.0177340

McGuire, A. L., Fisher, R., Cusenza, P., Hudson, K., Rothstein, M. A., McGraw, D., et al. (2008). Confidentiality, privacy, and security of genetic and genomic test information in electronic health records: points to consider. *Genet. Med.* 10, 495–499. doi: 10.1097/GIM.0b013e31817a8aaa

Meier, W., and Staffelbach, O. (1989). Fast correlation attacks on certain stream ciphers. *J. Cryptol.* 1, 159–176.

Mo, Y., Garone, E., Casavola, A., and Sinopoli, B. (2010). "False data injection attacks against state estimation in wireless sensor networks," in *49th IEEE Conference on Decision and Control (CDC)* (Atlanta, GA), 5967–5972.

Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Verezemska, O., Isbandi, M., et al. (2017). Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements. *Nucleic Acids Res.* 45, D446–D456. doi: 10.1093/nar/gkw992

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

Narayanan, A., and Shmatikov, V. (2008). "Robust de-anonymization of large sparse datasets," in *2008 IEEE Symposium on Security and Privacy (sp 2008)* (Oakland, CA), 111–125.

Newman, L. H. (2018). *The Worst Cybersecurity Breaches of 2018 so Far*. Available online at: https://www.wired.com/story/2018-worst-hacks-so-far/. (accessed July 9, 2018).

NMPDR (2019). *National Microbial Pathogen Data Resource*. Available online at: http://www.nmpdr.org (accessed March, 2019).

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189

Osborne, C. (2018). *US Hospital Pays $55,000 to Hackers After Ransomware Attack*. Available online at: https://www.zdnet.com/article/us-hospital-pays-55000-to-ransomware-operators/. (accessed January 9, 2018).

Papatheodorou, I., Fonseca, N. A., Keays, M., Tang, Y. A., Barrera, E., Bazant, W., et al. (2018). Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* 46, D246–D251. doi: 10.1093/nar/gkx1158

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2018). Cyberbiosecurity: from naive trust to risk awareness. *Trends Biotechnol.* 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012.

Pedro, H., Maheswari, U., Urban, M., Irvine, A. G., Cuzick, A., McDowall, M. D., et al. (2016). PhytoPath: an integrative resource for plant pathogen genomics. *Nucleic Acids Res.* 44, D688–D693. doi: 10.1093/nar/gkv1052

Pendleton, K. M., Erb-Downward, J. R., Bao, Y., Branton, W. R., Falkowski, N. R., Newton, D. W., et al. (2017). Rapid pathogen identification in bacterial pneumonia using real-time metagenomics. *Am. J. Res. Crit. Care Med.* 196, 1610–1612. doi: 10.1164/rccm.201703-0537LE

Pickett, B. E., Greer, D. S., Zhang, Y., Stewart, L., Zhou, L. W., Sun, G. Y., et al. (2012). Virus pathogen database and analysis resource (vipr): a comprehensive bioinformatics database and analysis resource for the coronavirus research community. *Viruses Basel* 4, 3209–3226. doi: 10.3390/v4113209

Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families database. *Nucleic Acids Res.* 40, D290–301. doi: 10.1093/nar/gkr1065

Ravan, R. R., Idris, N. B., and Mehrabani, Z. (2013). "A survey on querying encrypted data for database as a service," in *2013 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (Beijing), 14–18. doi: 10.1109/CyberC.2013.12

Samarati, P., and Sweeney, L. (1998). *Protecting Privacy When Disclosing Information: k-anonymity and its Enforcement Through Generalization and Suppression.* Technical report, technical report, SRI International.

Sandhu, R. S. (1992). "The typed access matrix model." in *Proceedings 1992 IEEE Computer Society Symposium on Research in Security and Privacy* (Oakland, CA), 122–136.

Sayers, S., Li, L., Ong, E., Deng, S., Fu, G., Lin, Y., et al. (2018). Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acids Res*. 47, D693–D700. doi: 10.1093/nar/gky999

Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G., et al. (2009). BioMart biological queries made easy. *BMC Genom.* 10:22. doi: 10.1186/1471-2164-10-22

Tahir, R., Huzaifa, M., Das, A., Ahmad, M., Gunter, C., Zaffar, F., et al. (2017). "Mining on someone else's dime: mitigating covert mining operations in clouds and enterprises," in *Research in Attacks, Intrusions, and Defenses. RAID 2017. Lecture Notes in Computer Science,* Vol. 10453, eds M. Dacier, M. Bailey, M. Polychronakis, and M. Antonakakis (Cham: Springer).

Tong, X., Xu, H., Zou, L., Cai, M., Xu, X., Zhao, Z., et al. (2017). High diversity of airborne fungi in the hospital environment as revealed by meta-sequencing-based microbiome analysis. *Sci. Rep.* 7:39606. doi: 10.1038/srep39606

Urban, M., Cuzick, A., Rutherford, K., Irvine, A., Pedro, H., Pant, R., et al. (2017). PHI-base: a new interface and further additions for the multi-species pathogen host interactions database. *Nucleic Acids Res.* 45, D604–D610. doi: 10.1093/nar/gkw1089

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., et al. (2017). Improvements to patric, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Res.* 45, D535–D542. doi: 10.1093/nar/gkw1017

Winnenburg, R. (2006). PHI-base: a new database for pathogen host interactions. *Nucleic Acids Res.* 34, D459–D464. doi: 10.1093/nar/gkj047

Xiang, Z., Tian, Y., and He, Y. (2007). PHIDIAS: a pathogen-host interaction data integration and analysis system. *Genome Biol.* 8:R150. doi: 10.1186/gb-2007-8-7-r150

Xu, L., Chen, L., Shah, N., Gao, Z., Lu, Y., and Shi, W. (2017). "Dl-bac: distributed ledger based access control for web applications," in *WWW '17 Companion Proceedings of the 26th International Conference on World Wide Web Companion*, (Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee), 1445–1450.

Xu, Y., Hunt, T., Kwon, Y., Georgiev, M., Shmatikov, V., and Witchel, E. (2016). "Earp: Principled storage, sharing, and protection for mobile apps," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)* (Santa Clara, CA: USENIX Association), 627–642.

Zhong, S., Yang, Z., and Wright, R. N. (2005). "Privacy-enhancing k-anonymization of customer data," in *Proceedings of the Twenty-fourth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '05 (New York, NY: ACM), 139–147. doi: 10.1145/1065167.1065185

Zmora, N., Zeevi, D., Korem, T., Segal, E., and Elinav, E. (2016). Taking it personally: personalized utilization of the human microbiome in health and disease. *Cell Host Microbe* 19, 12–20. doi: 10.1016/j.chom.2015.12.016

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Bioengineering and Biotechnology

REVIEW
published: 31 May 2019
doi: 10.3389/fbioe.2019.00116

Cyberbiosecurity for Biopharmaceutical Products

Jennifer L. Mantle[1], *Jayan Rammohan*[2†], *Eugenia F. Romantseva*[2†], *Joel T. Welch*[3], *Leah R. Kauffman*[2], *Jim McCarthy*[4], *John Schiel*[2], *Jeffrey C. Baker*[3], *Elizabeth A. Strychalski*[2], *Kelley C. Rogers*[2,5] and *Kelvin H. Lee*[1]*

[1] National Institute for Innovation in Manufacturing Biopharmaceuticals, Newark, DE, United States, [2] Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, United States, [3] Office of Biotechnology Products (OBP), Center for Drug Evaluation and Research (CDER), U.S. Food and Drug Administration, Silver Spring, MD, United States, [4] Information Technology Laboratory, National Cybersecurity Center of Excellence, National Institute of Standards and Technology, Gaithersburg, MD, United States, [5] Office of Advanced Manufacturing, National Institute of Standards and Technology, Gaithersburg, MD, United States

OPEN ACCESS

Edited by:
Randall Steven Murch,
Virginia Tech, United States

Reviewed by:
Calvin Chue,
U.S. Army Edgewood Chemical Biological Center (ECBC), United States
Maria José Saavedra,
Universidade de Trás os Montes e Alto Douro, Portugal

***Correspondence:**
Kelvin H. Lee
KHL@udel.edu

† These authors have contributed equally to this work

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *25 March 2019*
Accepted: *07 May 2019*
Published: *31 May 2019*

Citation:
Mantle JL, Rammohan J, Romantseva EF, Welch JT, Kauffman LR, McCarthy J, Schiel J, Baker JC, Strychalski EA, Rogers KC and Lee KH (2019) Cyberbiosecurity for Biopharmaceutical Products. Front. Bioeng. Biotechnol. 7:116. doi: 10.3389/fbioe.2019.00116

Cyberbiosecurity is an emerging discipline that addresses the unique vulnerabilities and threats that occur at the intersection of cyberspace and biotechnology. Advances in technology and manufacturing are increasing the relevance of cyberbiosecurity to the biopharmaceutical manufacturing community in the United States. Threats may be associated with the biopharmaceutical product itself or with the digital thread of manufacturing of biopharmaceuticals, including those that relate to supply chain and cyberphysical systems. Here, we offer an initial examination of these cyberbiosecurity threats as they stand today, as well as introductory steps toward paths for mitigation of cyberbiosecurity risk for a safer, more secure future.

Keywords: cyberbiosecurity, cybersecurity, biopharmaceutical manufacturing, engineering biology, cell therapy, gene therapy, supply chain

INTRODUCTION

Cyberbiosecurity is an emerging discipline encompassing vulnerabilities and corrective measures needed to address the unique risks existing at the intersection of cybertechnology and biotechnology. An early, inclusive definition of cyberbiosecurity is "understanding the vulnerabilities to unwanted surveillance, intrusions, and malicious and harmful activities which can occur within or at the interfaces of comingled life and medical sciences, cyber, cyber-physical, supply chain and infrastructure systems, and developing and instituting measures to prevent, protect against, mitigate, investigate, and attribute such threats as it pertains to security, competitiveness and resilience" (Murch et al., 2018).

To place context around the area of *cyberbiosecurity*, it is worth reviewing the established terms that contribute to this emerging discipline. *Cybersecurity* considers the security of digital information that is propagated and stored through networks of connected electronic devices (Lord, 2019). In general, *biosecurity* refers to the threat to living organisms and the environment due to exposures to biological agents, such as pathogens, whether occurring naturally or intentionally created (Institute of Medicine and National Research Council, 2006). A *cyber-biological interface* results when biological information is measured, monitored, or altered, and converted to digital information, or in the reverse, when digital information is used to manipulate a biological system. Similarly, a *cyber-physical interface* occurs when a physical mechanism is controlled or monitored by a digital means, such as the computer controlled mixing speed of a bioreactor. Importantly, cyber-physical interfaces may alter biological properties, blurring the lines of individualized definitions. Our intent in this publication is not to further refine the definition of cyberbiosecurity, as we believe that is best done through ongoing dialog within relevant stakeholder communities. Therefore, we rely on a working understanding of cyberbiosecurity as stated by (Peccoud et al., 2017), in referring to "the new risks emerging at the frontier between cyberspace and biology." For the purposes of this paper, we focus on cyberbiosecurity for the manufacture of biopharmaceuticals, to raise awareness of the existing risks that will be compounded through innovation in both the emerging types of biologically-manufactured therapies and the increasingly-automated processes used to develop and manufacture them.

The biopharmaceutical industry contributes nearly one trillion dollars to the U.S. economy, and has been highly successful in industrializing biotechnologies to produce biologic therapeutics (PhRMA, 2017). Biopharmaceutical products, or biologics, use engineered biological systems as platforms to manufacture therapeutic products to prevent or treat a variety of health conditions, such as cancer, diabetes, autoimmune disorders, and microbial infections. These products include vaccines, traditional protein therapeutics, such as monoclonal antibodies, as well as emerging biotechnologies, such as cell and gene therapies.

Although the processes differ in how various classes of therapeutics are manufactured, in each process, information flows repeatedly between biological information (i.e., genetic) and cyber (i.e., digital) information. Securing this information flow through thoughtful assessment of vulnerabilities and threats for biopharmaceutical manufacturing is critical for public health, economic security, and national security. The focus of this publication is to illuminate these vulnerabilities and threats to encourage the broad stakeholder community to work toward the development of appropriate risk mitigation strategies, both for the current state-of-the-art and for the emerging technologies that represent the future state of the industry. Novel threats to the security of biological and related information along interfaces relevant to human health and manufacturing processes will continue to emerge as innovation progresses.

The interface of biological and digital information in biomanufacturing creates two primary concerns in evaluating cyberbiosecurity vulnerabilities, that recur throughout multiple processes in the end-to-end workflow (see Figure 1, Peccoud et al., 2017). The first concern is the nature of the biological manufacturing platform, as information contained in biological systems is subject to both evolution and context in ways that may not be well-understood or predictable. The variation that biological systems introduce in manufacturing presents risks for product consistency. The industry has developed extensive bioprocess control strategies and release testing to mitigate risks for established classes of biotherapeutics to ensure consistent product with minimal lot-to-lot variability. However, this biological variation presents challenges for innovating flexible scaling of existing large-batch processes. The issue of inherent biological variation is a critical challenge in the manufacture of emerging classes of gene and cellular therapies where typical small-batch manufacturing across a wider diversity of product types precludes the reliance on large historical data sets to allow identification of subtle process deviation. For these small-batch products, subtle genetic deviation during cellular expansion steps may be magnified *in vivo* due to differences between the host and the patient.

The second area of concern is the integrity of the data associated with the biopharmaceutical manufacturing process, including data related to supply chain and cyberphysical systems. Biopharmaceutical manufacturers are complex organizations that rely on technology as part of daily operations to tightly monitor and control biopharmaceutical production processes. The notion of a digital thread, which refers to data that follows a product and informs decisions throughout its life cycle, can be applied to the biopharmaceutical industry (Wang, 2018). The digital thread of the manufacturing of biopharmaceuticals includes data that support the development and scale up of the manufacturing process, clinical data, post-approval data, and the equipment used to manufacture the product. As the number of interconnected devices and systems that inform digital threads increases, cybersecurity vulnerability increases, because one vulnerable device can result in a threat that compromises a single point, or an entire process, system, or supply chain. Further, as a result of greater dependence on automation and decentralized manufacturing, the security of information transfer from site to site is critical to ensure the efficacy of the production process. While many cybersecurity concerns related to biopharmaceutical processes can be mitigated by existing best practices, standards, and regulations, the additional complexities at the cyber-biological interfaces during biopharmaceutical manufacturing processes, described below, warrant further examination.

The relevant stakeholder communities should establish a means of identifying and assessing the potential new vulnerabilities and threats, toward the development of effective risk mitigation strategies. For example, the NIST Framework for Improving Critical Infrastructure Cybersecurity is a voluntary, standards-based approach for identifying and protecting assets and systems, and detecting, responding to, and recovering from cyber intrusions (NIST, 2018). While the framework was originally developed for critical infrastructure systems where it has been widely adopted since its introduction in 2014, its focus on business drivers for cybersecurity risk assessment and practices makes it broadly applicable to many industries.

To further encourage the community's consideration of cyberbiosecurity vulnerabilities and mitigations, we include insights into the development of current cybersecurity best practices and guidance for medical devices, as a useful model for the path forward for a best-practices risk-mitigation framework for cyberbiosecurity for biopharmaceutical manufacturing. It is our hope that current biopharmaceutical industry practices can inform risk-mitigation for emerging classes of biotherapeutics and innovative production platforms for established classes of biotherapeutics. Current practices may also illuminate parallel considerations related to cyberbiosecurity in other biomanufacturing sectors and applications, such as synthetic biology approaches to the production of commodity chemicals and biofuels.

RISKS ASSOCIATED WITH THE BIOLOGICAL MANUFACTURING PLATFORM IN BIOPHARMACEUTICAL MANUFACTURING WORKFLOWS

While best practices for cybersecurity apply to biopharmaceutical manufacturing, biological systems present unique vulnerabilities in production processes. Cyberbiosecurity vulnerabilities may be considered with regard both to using an engineered biological system as the manufacturing platform, as is the case for protein therapeutics, and for products that are themselves an engineered biological system, as for cellular therapies. The dynamic nature of genetic information that aids survival in natural environments poses challenges in engineering and manufacturing settings. For example, some change in the genetic information of a cell population is unavoidable during expansion and growth in a bioreactor, so biomanufacturing processes must contend with heterogeneous populations of cells that may yield a heterogeneous product, whether biomolecular or cellular. The ability of biological systems to alter the content and expression of their genetic information presents significant complexity for biopharmaceutical manufacturing unique to those posed by cyber-systems that must be considered in strategies for cyberbiosecurity risk mitigation.

Challenges of Genetic Information

Two fundamental distinctions between digital and biological information are relevant in considering the cyber-biological interface during the end-to-end biopharmaceutical manufacturing process. First, genetic information evolves naturally when replicated. Mechanisms that drive natural changes in DNA sequence include mutation, recombination, horizontal gene transfer, and others. Second, the expression of this information can change depending on how an organism senses and responds to its environment. This dependence on context, which encompasses all aspects of the system in which the genetic information exists, cannot always be predicted. The same sequence of DNA may have dramatically different consequences for function depending on surrounding DNA sequences, intra- and inter-molecular interactions within the cell, and extracellular conditions. Thus, the impact of changes, whether due to natural "drift" or through malicious introduction, is difficult to predict, detect, and mitigate.

Protein Therapeutics

State-of-the-art biomanufacturing of protein therapeutics uses engineered mammalian cells as the manufacturing platform. One notable example is Chinese hamster ovary (CHO) cells used as the host cell system (Jayapal et al., 2007). To better assess potential vulnerabilities at the cyber-biological interface in this process, we consider the flow of genetic information in a typical biomanufacturing workflow.

The security of the genetic information at the cyber-biological interface is assured initially through the integrity of the nucleic acid used to transfect a cell line. Programmable DNA synthesizers and sequencers specify and confirm the DNA sequence that is then stably transfected into host cells for cell line development. This process effectively transfers digital information into a "genetic thread" that parallels the digital thread of the manufacturing process. A selection of clonal cells with desired phenotypes for yield and stability are then passaged under defined conditions to produce master cell banks, which are passaged further to produce working and production cell banks. Throughout these workflows, consistent cell culture expansion protocols are used to achieve consistent context for the genetic information, with the intent of minimizing natural mutations. Contextual security of the genetic information during production is also maximized through well-defined process control strategies. This context includes bioreactor growth conditions, such as feeding strategy, dissolved oxygen concentration, gas flow, sparge rates, pH, and temperature. Cell populations that exhibit genetic instability during bioreactor growth are identified through deviations from established process parameters, so that processes can be aborted at early stages, and there is no risk to product quality. Genetic stability across the expanded cell populations is also monitored for transgene sequence and copy number, including the testing of post-production cell banks to ensure data across the full thread of genetic information. As the natural evolution of the cells during expansion cannot be reversed, the security of the master cell banks is critical to ensure the consistency of the product through its lifecycle, and redundancies are built into storage strategies to guard against any single failure mode.

At the state of the art, the industry is mitigating risks associated with the uncertainty in product safety profiles due to natural variation or contamination in the biological system, through extensive control and quality assurance strategies, following established best practices and rigorous regulatory guidance. Furthermore, as facility access is currently managed to ensure both protection of trade secrets and compliance with current U.S. FDA Good Manufacturing Practices regulations, it is difficult to imagine scenarios where malicious or adventitious acts on bioprocess workflows would go undetected for established manufacturing facilities producing protein therapeutics through large batch processes. However, a malicious intrusion increases uncertainty at the cyber-biological interface and could trigger batch losses, with significant economic impacts for the industry and could potentially result in drug shortages (Castellanos and Janofsky, 2018).

During the production of protein therapeutics, cyberbiosecurity vulnerabilities exist at each point where genetic information is stored, expressed, replicated, or monitored through cyber or cyber-physical systems. A simple example is the storage of master cell banks in a freezer with networked alarm and temperature monitoring systems, where failure in the network can introduce uncertainty in the viability of the master cell bank. A more malicious variation of this simple scenario is a cyber-intrusion that corrupts the digital record that documents the storage conditions for the master cell bank. In both cases, the uncertainty of the cells' viability presents a vulnerability, even if the actual impact on the stored cells was negligible.

A more complex example of a dynamic cyber-biological interface is a perfusion bioreactor. In this process, flow rates of media into the reactor and biomass removal out of the reactor are balanced to maintain a desired cell density within the bioreactor. The cell density is optimized for process yield and growth rate is controlled through parameters such as nutrient limitation (Bielser et al., 2018). The cyberphysical components of the system control media and biomass flow rates, which in turn constrain cellular growth rate and product yield. Thus, the vulnerabilities associated with the cyberphysical control system propagate into vulnerabilities in the biological output of the process.

As typical workflows for the production of protein therapeutics are fully established and industrialized, many of the risks are mitigated by current manufacturing practices. However, this discussion is intended to prompt a systematic evaluation of vulnerabilities and threats at the cyber-biological interfaces for these processes, both to reduce remaining vulnerabilities to malicious acts, and to inform risk-mitigation strategies for less-industrialized manufacturing workflows.

Emerging Classes of Biologic Therapies

Increasingly, engineered cells are themselves the therapeutic product, rather than simply serving as the biomanufacturing platform. For example, CAR-T cells (Androulla and Lefkothea, 2018) and engineered microbiome modulators (Garber, 2015) are members of a growing category of existing living therapeutics enabled by engineering biology methods. For these living therapeutics, as well as for *in vivo* gene therapies, the flow of genetic information occurs in both the production for the therapeutic agent, and within the patient. Each of the biosecurity considerations for protein therapeutics applies to living therapeutic modalities, but protein therapeutics benefit from decades of experience in production, as well as testing of product lot releases to identify, in principle, any relevant deviations in the flow of genetic information. Aside from unwanted physicochemical degradation, protein therapeutics cannot alter their own properties or respond to environmental context. Established process controls and quality assurances in protein therapeutic biomanufacturing should be adapted to address the emerging cyberbiosecurity needs of emerging novel modalities. However, emerging product modalities such as cellular and gene therapies convey alterations in genetic information that are intended to become self-replicating and expressed *in vivo*. These emerging therapies therefore pose additional safety concerns for patients that warrant further cyberbiosecurity evaluation of their manufacturing workflows, as well as pharmacovigilance at the patient level to monitor the integrity of the transferred genetic code.

Future Therapeutic Modalities

Engineered cells from all domains of life, including prokaryotes, eukaryotes, and archaea, as well as synthetic systems, such as cell-free systems, may offer potential biomanufacturing platforms and products in industrial workflows. The ongoing evolution of biotechnology fueled by increasingly automated DNA design, read, and write capabilities, along with facile gene-editing platforms, such as CRISPR, TALENs, and zinc-finger nucleases will continue to create new cyber-biological interfaces and additional risks for both biosecurity and biosafety.

Proof-of-concept exists for designing genetic circuits that can be used to encode logic in bacteria and enable them to perform clinically-relevant functions (Brophy and Voigt, 2014). In principle, cells could be engineered using genetic circuits to treat a wide range of pathologies, including but not limited to autoimmune diseases, cancer, and viral infections (Piñero-Lambea et al., 2015; Xie and Fussenegger, 2018). Computational methods that leverage principles from electronic design automation have been employed for the design and optimization of these genetic circuits (Nielsen et al., 2016). Genetic circuit design software, such as that offered by Teselagen, can automatically generate machine-readable synthesis instructions. Any processes similar to these, which involve the transfer of information between digital and biological forms, are potential points of vulnerability. While current biomanufacturing processes may be difficult to disrupt without detection, fully automated, distributed and "on-demand" biomanufacturing workflows of the future may make it possible to use malicious cyber-intrusions to corrupt the design, reading, and writing of DNA sequences to produce pathogenic, self-replicating entities that pose both biosecurity and biosafety hazards. Although these risks are still emerging, the rapid pace of innovation dictates that it is not too early to consider the cyberbiosecurity implications of such capabilities. The National Academies of Sciences, Engineering, and Medicine have recently assembled a committee to consider strategies on Safeguarding the Bioeconomy that is expected to contain an analysis of the unique elements of the biotechnology economy that will consider whether specific features of the bioeconomy may require innovative cybersecurity solutions.

Future Cyber-Biological Interfaces Enabled by Artificial Intelligence

Digital data may become increasingly similar to biological data, in that digital data may become more dynamic and dependent on its context, especially considering the expanding capabilities of artificial intelligence (AI) and the increasingly widespread implementation of machine learning algorithms. Looking forward, computers and biology in the same control loop is an emerging area that could introduce new cyberbiosecurity vulnerabilities as AI and machine learning become more mainstream. While current AI capabilities are mostly associated with passive learning, systems capable of active learning and neural networks are currently being developed for many different applications (Murphy, 2011; Lou et al., 2014; Angermueller et al., 2016; Jamali et al., 2016; Feltes et al., 2018). As artificial intelligence finds increasing application in biomanufacturing and transitions from completely dependent to semiautonomous to completely autonomous, a full assessment of vulnerabilities and threats should include strategies for mitigation. With each advance, cybersecurity and cyberbiosecurity may more fully approach a single, unified discipline.

CYBERBIOSECURITY, PROCESS CONTROL AND QUALITY/RISK MANAGEMENT

Biopharmaceutical manufacturing relies on complex technology as part of daily operations to tightly monitor and control biological production processes. Many of the failure modes in biopharmaceutical manufacturing are foundationally similar to those of other manufacturing modalities, and existing best practices in cybersecurity should be incorporated to mitigate those risks. The complexity of the digital thread arising from the biological component of the manufacturing process for biopharmaceuticals introduces additional risks that can impact product quality. We, therefore, propose that cyberbiosecurity should also be considered as a failure mode in the development of a manufacturing control strategy, and in maintenance of the validated state.

Physicochemical and biophysical data related to a biopharmaceutical product, for example, is generated throughout its lifecycle, detailing early generation products, reference material qualification, stability testing, and release strategies. Biological License Applications summarize this data through submission of a common technical document to regulatory authorities. The data originator must safeguard both raw and processed forms of this data for extended periods, typically years, for trending, re-evaluation, and comparison to support future comparability studies.

As mentioned in the above section, the aftermath of a cyberbiosecurity failure can have a significant impact on supply of medicines and on patient health. For example, many biopharmaceutical products are high-potency, low-volume operations, with a year or more of inventory generated in a single lot. A failure in such a manufacturing process would dangerously deplete the supply of that product. Furthermore, many biopharmaceutical processes contain non-compressible timelines (e.g., expansion cultures or hydrodynamic limitations), so timely recovery from a cyberbiosecurity failure could be difficult, especially in a high-utilization, multi-product plant. Patients that rely on biopharmaceuticals can be especially impacted by shortages or recalls because it is not uncommon for biopharmaceutical products to be presented through extended courses of therapy that have negative clinical consequences if interrupted.

Pharmaceutical Quality Management Systems

Pharmaceutical Quality Management Systems (QMS) are implemented to deliver products with appropriate quality attributes, establish and maintain a state of control, and facilitate continual improvement in manufacturing processes. By necessity, a QMS assumes that valid monitoring and assessment of the process are in place. A cyberbiosecurity breach has the potential to "break" an integrated QMS. If fundamental QMS activities, such as in-process and finished product analysis, inventory management, document management, change control, lot disposition, corrective actions, and preventative actions, were compromised by a cyberbiosecurity breach, biopharmaceutical manufacturing operations would have to either be shut down or subject to detailed, manual review, and assessment. This practice at best increases costs and human-sourced variation, and at worst compromises the quality of the product produced.

A QMS anticipates and detects special cause variation in the context of common cause variation. A cyberbiosecurity failure could present itself as an unanticipated or undetected special cause failure (e.g., an adventitious or malicious alteration or contamination of the data stream), or could cloud understanding of common cause variation (e.g., system decay or continuous improvement of operations). In particular, undetected cyberbiosecurity "contaminations" could be particularly worrisome. An undetected cyberbiosecurity failure could manifest in, for example, incorrect test results or expiry dates, incorrect process control loops and algorithms, inappropriate conduct of maintenance in the plant, or even disruption through presentation of false failures during inspection by regulators. For these reasons, assessment of cyberbiosecurity vulnerabilities should be built into a lifecycle control maintenance plan assuring the validated state. Different manufacturing processes may require different risk-based cyberbiosecurity measures to address different threats and vulnerabilities, however they should all be framed by the QMS.

Continuous improvement in biopharmaceutical manufacturing is predicated upon comparability exercises. Cyberbiosecurity failures that compromise the integrity of comparability can prevent continuous improvement and deployment of new technologies or manufacturing sites.

Manufacturing Process Control and Product Quality

Manufacture of traditional biopharmaceutical products, such as protein therapeutics, has a high level of residual uncertainty, making this type of manufacturing particularly vulnerable to cyberbiosecurity failure modes. Increasingly, biopharmaceutical manufacturers are employing a greater dependence on process analytical technologies, automation, and distributed and integrated control systems, with fewer manual interventions. This shift decreases human factor-related failure, but increases the likelihood of cyberbiosecurity-related failures for biopharmaceuticals.

Because engineered biological systems are used as the manufacturing platform, control of the product is a function of control and evaluation of multiple critical quality attributes (CQAs) and process parameters rather than direct measurement of clinically relevant mechanistic functions. Process control strategies monitor common and special cause variability, sort variations into relevant (signal) and indeterminate (noise), and trigger corrective actions. The acts of monitoring, sorting, and communicating corrective actions are vulnerable to cyberbiosecurity threats. These failure modes can lead to special cause errors, which can subsequently lead to false or misleading signals, or undetected or uncommunicated process failures.

As those in process development increase use of process analytical technology (e.g., on-line/at-line testing) and move toward real-time release, there is less opportunity for detection and mitigation of a cyberbiosecurity breach. For example, processes depend more upon validated clearance of process-specific contaminants (e.g., DNA, viruses, host cell proteins, residual solvent, etc.) rather than lot-to-lot testing. A compromise to the validated processing envelope in the form of a cyberbiosecurity breach could impact product quality and safety because the assumption supporting clearance established during process validation would no longer be valid. A shift in process control toward real-time release could increase the possible impact of a cyberbiosecurity failure compared to lot-to-lot release testing. This is not to say that real time release practices should be avoided but rather that dynamic risk assessment modeling is crucial to understanding these advanced control strategies.

Manufacturing Supply Chain Considerations

Biopharmaceutical manufacturing frequently uses reagents or materials with few alternative vendors. The risk and impact of a cyberbiosecurity failure within the supply chain or at a key vendor could have an unanticipated, negative impact on the assurance of a consistent supply of high-quality biopharmaceuticals. A second supply chain consideration is for the biopharmaceutical product itself. These products are often sterile parenterals with cold chain conformance requirements. Indirect adventitious or malicious cyberbiosecurity attacks to maintenance of sterile operations or to the cold chain could lead to loss of product or, worse, could compromise patient safety or efficacy.

Cybersecurity for Medical Devices as a Model for Developing Risk Evaluation and Mitigation Strategies for Cybersecurity Vulnerabilities in Biopharmaceutical Manufacturing

The medical device industry faced a similar challenge, as medical devices create a cyber-biological interface with direct patient impact. As devices become increasingly interconnected, cybersecurity concerns for medical devices, such as device access and security of information and data, drove community engagement to develop best practices to address these concerns. Cybersecurity specifically refers to the protection of computer systems, including hardware, software, and data, from unauthorized access, theft, damage, disruption or misdirection. A medical device itself has hardware, software, and data that could potentially be compromised after a cybersecurity attack. The community engaged with the FDA Center for Devices and Radiological Health (CDRH) to systematically evaluate risks at all points in the device life cycle and then to develop best practices to mitigate these risks. As a result of these efforts, CDRH has released three Guidance for Industry documents [FDA., 2014, 2016, 2018 (draft)], and hosted four public workshops where discussion of medical device technology, device regulation, policy gaps, and best practices was welcomed. Similarly, community engagement between all stakeholders including industry and regulators, could lead to the development of best practices for cyberbiosecurty in the biopharmaceutical industry.

CONCLUSIONS

Biopharmaceutical products have had a substantial positive impact on public health. With the increasing digitalization of information related to such products and how they are manufactured, it becomes important to consider potential impacts from cyberbiosecurity-related threats. Detected intrusions will trigger the need for investigation and mitigation within a robust quality management system. Among the potential impacts are:

- Economic loss to the industry due to a manufacturing process out of specifications, poor product quality, or loss of confidence in the integrity of the process.
- Patient and public health impacts due to ineffective, dangerous, or lost production batches, most notably for autologous therapies, such as CAR-T.
- Exposure of employees to harmful agents, for example, through the deliberate introduction of a pathogen into manufacturing process.
- Inability to respond rapidly to emergent public health threats.

Therefore, analysis is warranted to identify and mitigate the unique cyberbiosecurity risks and failure modes in the biopharmaceutical industries. Current best practices from industrial manufacturing and state-of-the-art cybersecurity could serve as a starting point to safeguard and mitigate against cyberbiosecurity threats to biomanufacturing.

Given the importance of the issues raised by cyberbiosecurity risks, ecosystem-wide coordination and communication to develop a more comprehensive understanding of the field as well as appropriate mitigation strategies are needed. One possible path forward may be to explore the use of NIST's Framework for Improving Critical Infrastructure Cybersecurity to manage risks introduced by vulnerabilities and threats unique to biological systems. The framework could potentially be adapted or profiled with input from stakeholders to include relevant standards, guidelines, and best practices to manage cyberbiosecurity risks for biomanufacturing organizations of all scales. The framework could allow businesses and organizations to develop their own unique profile to address risk appetite, mission priority, budget, and resource constraints within the scope of their requirements, objectives, and desired outcomes. A follow-on publication from NIST provides a manufacturing-specific roadmap for reducing cybersecurity risk that may provide additional guidance to the biomanufacturing community (Stouffer et al., 2017).

Cyberbiosecurity concerns should be a part of modern, risk-based, quality management systems and should be considered in the development and maintenance of process control strategies throughout the product life cycle. Education and awareness of existing best practices for cybersecurity of manufacturing systems is essential for personnel involved in any stage of these processes. Creating standard practices to fully incorporate cyberbiosecurity awareness into every stage of the biomanufacturing process can lead to a more secure supply of safe, life-saving medicines, ultimately improving lives through a healthy society, and strong economy.

AUTHOR CONTRIBUTIONS

JLM, KR, and KL contributed to all sections of the manuscript. JW and JB led the writing of the section on biopharmaceutical manufacturing and made comments and edits on other sections. JR, ER, LK, JM, JS, and ES led the writing of the section on biopharmaceutical products and made comments and edits on other sections.

FUNDING

JM and KL performed this work under financial assistance award 70NANB17H002 from U.S. Department of Commerce, National Institute of Standards and Technology.

REFERENCES

Androulla, M. N., and Lefkothea, P. C. (2018). CAR T-cell therapy: a new era in cancer immunotherapy. *Curr. Pharm. Biotechnol.* 19, 5–18. doi: 10.2174/1389201019666180418095526

Angermueller, C., Pärnamaa, T., Parts, L., and Stegle, O. (2016). Deep learning for computational biology. *Mol. Syst. Biol.* 12:878. doi: 10.15252/msb.20156651

Bielser, J.-M., Wolf, M., Souquet, J., Broly, H., and Morbidelli, M. (2018). Perfusion mammalian cell culture for recombinant protein manufacturing – a critical review. *Biotechnol. Adv.* 36, 1328–1340. doi: 10.1016/j.biotechadv.2018.04.011

Brophy, J. A. N., and Voigt, C. A. (2014). Principles of genetic circuit design. *Nat. Methods* 11, 508–520. doi: 10.1038/nmeth.2926

Castellanos, S., and Janofsky, A. (2018). *One Year After notpetya Cyberattack, Firms Wrestle With Recovery Costs*. The Wall Street Journal. Available online at: https://www.wsj.com/articles/one-year-after-notpetya-companies-still-wrestle-with-financial-impacts-1530095906

FDA. (2014). *Content of Premarket Submissions for Management of Cybersecurity in Medical Devices Guidance for Industry and Food and Drug Administration Staff*. FDA Guidance, 6.

FDA. (2016). *Postmarket Management of Cybersecurity in Medical Devices. Guidance for Industry and Food and Drug Administration Staff*. Food Drug Administration. 1–30. Available online at: https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm482022.pdf

FDA. (2018). *Content of Premarket Submission Management of Cybersecurity in Medical Devices: Draft Guidance for Industry and Food and Drug Administration Staff. (DRAFT)*. Available online at: https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM623529.pdf (accessed March 14, 2019).

Feltes, B. C., Grisci, B. I., Poloni, J., de, F., and Dorn, M. (2018). Perspectives and applications of machine learning for evolutionary developmental biology. *Mol. Omi.* 14, 289–306. doi: 10.1039/C8MO00111A

Garber, K. (2015). Drugging the gut microbiome. *Nat. Biotechnol.* 33, 228–231. doi: 10.1038/nbt.3161

Institute of Medicine and National Research Council. (2006). *Globalization, Biosecurity, and the Future of the Life Sciences*. Washington, DC: The National Academies Press

Jamali, A. A., Ferdousi, R., Razzaghi, S., Li, J., Safdari, R., and Ebrahimie, E. (2016). DrugMiner: comparative analysis of machine learning algorithms for prediction of potential druggable proteins. *Drug Discov. Today* 21, 718–724. doi: 10.1016/j.drudis.2016.01.007

Jayapal, K. P., Wlaschin, K. F., Hu, W. S., and Yap, M. G. S. (2007). Recombinant protein therapeutics from CHO Cells - 20 years and counting. *Chem. Eng. Prog.* 103, 40–47.

Lord, N. (2019). *What is Cyber Security? Definition, Best Practices and More*. DataInsider. Available online at: https://digitalguardian.com/blog/what-cyber-security

Lou, X., Schiegg, M., and Hamprecht, F. A. (2014). Active structured learning for cell tracking: algorithm, framework, and usability. *IEEE Trans. Med. Imaging* 33, 849–860. doi: 10.1109/TMI.2013.2296937

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

Murphy, R. F. (2011). An active role for machine learning in drug development. *Nat Chem Biol.* 7, 327–330. doi: 10.1038/nchembio.576

Nielsen, A. A. K., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., et al. (2016). Genetic circuit design automation. *Science* 352:aac7341. doi: 10.1126/science.aac7341

NIST. (2018). *Framework for Improving Critical Infrastructure Cybersecurity*. Available online at: https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.04162018.pdf (accessed March 14, 2019).

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2017). Cyberbiosecurity: from naïve trust to risk awareness. *Trends Biotechnol.* 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012

PhRMA. (2017). *The Economic Impact of the U.S. Biopharmaceutical Industry: 2015 National and State Estimates*. Available online at: http://phrma-docs.phrma.org/files/dmfile/PhRMA_GoBoldly_Economic_Impact.pdf (accessed March 14, 2019).

Piñero-Lambea, C., Ruano-Gallego, D., and Fernández, L. Á. (2015). Engineered bacteria as therapeutic agents. *Curr. Opin. Biotechnol.* 35, 94–102. doi: 10.1016/j.copbio.2015.05.004

Stouffer, K., Zimmerman, T., Tang, C., Lubell, J., Cichonski, J., and McCarthy, J. (2017). *Cybersecurity Framework Manufacturing Profile Cybersecurity*. Available online at: https://nvlpubs.nist.gov/nistpubs/ir/2017/NIST.IR.8183.pdf (accessed March 14, 2019).

Wang, B. (2018). The future of manufacturing: a new perspective. *Engineering* 4, 722–728. doi: 10.1016/j.eng.2018.07.020

Xie, M., and Fussenegger, M. (2018). Designing cell function: assembly of synthetic gene circuits for cell biology applications. *Nat. Rev. Mol. Cell Biol.* 19, 507–525. doi: 10.1038/s41580-018-0024-z

Disclosure: This publication reflects the views of the author and should not be construed to represent FDA's views or policies. Certain commercial entities, equipment, products, or materials may be identified by name in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply special status or relationship with NIST or recommendation or endorsement by NIST; neither is it intended to imply that the entities, equipment, products, or materials are necessarily the best available for the purpose.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Bioengineering and Biotechnology

PERSPECTIVE
published: 04 September 2019
doi: 10.3389/fbioe.2019.00210

Cyberbiosecurity in Advanced Manufacturing Models

*Donovan Guttieres, Shannon Stewart, Jacqueline Wolfrum and Stacy L. Springs**

Center for Biomedical Innovation, Massachusetts Institute of Technology, Cambridge, MA, United States

Cybersecurity for the production of safe and effective biopharmaceuticals requires the attention of multiple stakeholders, including industry, governments, and healthcare providers. Cyberbiosecurity breaches could directly impact patients, from compromised data privacy to disruptions in production that jeopardize global pandemic response. Maintaining cybersecurity in the modern economy, where advanced manufacturing technologies and digital strategies are becoming the norm, is a significant challenge. Here, we highlight vulnerabilities in present and future biomanufacturing paradigms given the dependence of this industry sector on proprietary intellectual property, cyber-physical systems, and government-regulated production environments, as well as movement toward advanced manufacturing models. Specifically, we (1) present an analysis of digital information flow in a typical biopharmaceutical manufacturing value chain; (2) consider the potential cyberbiosecurity risks that might emerge from advanced manufacturing models such as continuous and distributed systems; and (3) provide recommendations for risk mitigation. While advanced manufacturing models hold the potential for reducing costs and increasing access to more personalized therapies, the evolving landscape of the biopharmaceutical enterprise has led to growing concerns over potential cyber attacks. Gaining better foresight on potential risks is key for implementing proactive defensive principles, framing new developments, and establishing a permanent security culture that adapts to new challenges while maintaining the transparency required for regulated production of safe and effective medicines.

Keywords: cybersecurity, biomanufacturing, distributed manufacturing, bioprocess risks, cyber-physical systems, risk mitigation

OPEN ACCESS

Edited by:
Diane DiEuliis,
National Defense University,
United States

Reviewed by:
Paula Alexandra Oliveira,
University of Trás-os-Montes and
Alto Douro, Portugal
Kelvin H. Lee,
University of Delaware, United States

****Correspondence:***
Stacy L. Springs
ssprings@mit.edu

Specialty section:
This article was submitted to
Biosafety and Biosecurity,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

***Received:** 31 January 2019*
***Accepted:** 19 August 2019*
***Published:** 04 September 2019*

Citation:
Guttieres D, Stewart S, Wolfrum J and
Springs SL (2019) Cyberbiosecurity in
Advanced Manufacturing Models.
Front. Bioeng. Biotechnol. 7:210.
doi: 10.3389/fbioe.2019.00210

INTRODUCTION

Cybersecurity attacks and data breaches are a matter of when, not if, with companies in all sectors and of all sizes vulnerable. Between 2014 and 2015, the Federal Bureau of Investigation (FBI) reported a 53% increase in incidents of industrial or economic espionage targeted at the U.S (Barrett, 2015). In the healthcare industry, data collected by the Department of Health and Human Services shows a 10% increase in the number of reported incidents each year since 2010 (U S Department of Health and Human Services Office for CivilRights, 2019). In 2017, this industry accounted for 18% of all data breaches, with 63% of incidents caused by criminal or malicious activity.

More recently, cyber extortionists have targeted hospital IT systems, successfully extracting thousands of dollars in ransoms because of the critical and often time-sensitive nature of the information (Osborne, 2018). In one example, the WannaCry ransomware affected hospitals within

UK's National Health Service, leading to 6,912 medical appointment cancellations and 1,220 pieces of IT-connected diagnostic equipment infected, largely due to unpatched or unsupported Windows operating systems (National Audit Office, 2018). Increasing concerns over the potential threat cyberattacks can have on acquiring access to and control of medical devices, especially if digitally connected (e.g., insulin pumps) or on a hospital network (e.g., radiologic imaging equipment), has led the FDA to release pre- and post-market guidance to reduce cyber-related risks (U. S. Federal Drug Administration, 2018). Spaces such as the Biohacking Village (https://www.villageb.io/) encourage dialogue between medical device and cybersecurity professionals, but should also consider biomanufacturing-related vulnerabilities.

NEW RISKS WITHIN THE GROWING BIOECONOMY

The bioeconomy has become a principal driver of national GDP (National Academies of Sciences Engineering and Medicine, 2015). Nevertheless, there is insufficient attention and collection of data on the risk of cyberattacks targeting organizations that manufacture life-saving or -extending biologic medicines, such as vaccines, recombinant proteins, monoclonal antibodies, and advanced therapy medicinal products (ATMPs). Whether the biotechnology industry is deliberately targeted or collateral damage in cyber warfare, the effects could be severe given the high-value products and data involved.

The intellectual property, manufacturing processes, regulatory requirements and sophisticated cyber-physical systems involved in the production of biologic therapies may be particularly vulnerable to three major forms of cyberattacks: sabotage (deliberate and malicious acts that damage digital or physical infrastructure), corporate espionage (gaining access to sensitive information to attain advantage over an adversary), and crime/extortion (encrypting files with a ransom note asking for remuneration for their return) (Morag, 2014). Examples of each have been reported across the biotechnology industry (Panda Security., 2017; Sackner-Bernstein, 2017; Symantec., 2018). Despite differences in these modes of cyberattack, their mechanisms can often be similar (e.g., phishing attacks, malware, encryption blind spots, cloud-based threats, negligence, and poor institutional knowledge of risks). In all their forms, cybersecurity incidents raise serious concern for the biopharmaceutical industry, government, regulators, health service providers and ultimately patients.

The formalization of cyber*bio*security, at the nexus of cybersecurity, cyber-physical security and biosecurity as applied to biological and biomedical-based systems, provides insight into the unique risks present in the biotechnology industry (Murch et al., 2018). More specifically, biopharmaceutical companies employ cyber-physical systems across a range of functions: raw materials sourcing, cell line development and optimization, upstream and downstream process development, manufacturing, validation studies, clinical trials, supply chain management of products, post-market drug safety monitoring, and interfacing with health providers. Process control strategies increasingly collect and use data to ensure that manufacturing processes meet product quality standards. As part of advanced manufacturing approaches, various tools (e.g., internet-of-things, artificial intelligence) are allowing for more responsive control to optimize for reproducibility, quality, safety and supply (Helu and Hedberg, 2015; Zhong et al., 2017). However, in-line, at-line or remote data monitoring can also increase vulnerability to cyberattacks given the increasing reliance on digital and automated control systems (Babiceanu and Seker, 2016).

KNOWN CYBERSECURITY RISKS POINT TO VULNERABILITIES IN BIOMANUFACTURING

U.S. biopharmaceutical companies together spend nearly $160 billion each year on R&D, and their accumulated intellectual property (IP) is likely worth trillions of USD (Research America., 2016). An advanced, persistent attack could allow corporate rivals to steal internal communications, IP related to the product or process, and facility monitoring data to gain a competitive advantage. A malware program called Dragonfly specifically targets cyber-physical systems used in pharmaceutical manufacturing equipment, stealing trade and manufacturing secrets as a form of corporate espionage (Carman, 2014). Some have suggested that Dragonfly could also be used for physical sabotage in the future (Symantec Security Response., 2014). Pharmaceutical companies hold patient data related to clinical trials and disease management in their corporate networks. Since the data is both highly sensitive personal information and regulated, breaches can both incur large fines and damage a firm's reputation. Assessing emerging cybersecurity risks across the biopharmaceutical industry is especially important and timely as many companies work to establish digital strategies and data lakes that serve as repositories of data from across

Case Example - Merck & Co.: In June of 2017, the biopharmaceutical company Merck & Co. was affected by the malicious worm NotPetya (Erman and Finkle, 2017). The worm was based on ransomware, Petya, but it had been modified so that it was unable to revert its changes, resulting in the permanent encryption of data (Goodin, 2017). Since the malware affected computer systems that are used to control Merck's manufacturing process, the attack resulted in shortages of the Gardasil vaccine and may have contributed to stock-outs of the Hepatitis B vaccine. The incident led Merck to borrow $240 million worth of Gardasil vaccine from the Center for Disease Control's stockpile, with a total estimated cost of the cyberattack close to $1 billion (United States Securities and Exchange Commission, 2018). In February 2018, the US and UK publicly attributed the attack to Russia (Marsh, 2018). Since there is no evidence to believe that Merck had been deliberately targeted, it is easy to imagine a more tailored or intentional cyberattack causing even more damage to biomanufacturing activities. Given the low number of reported cases of cyberattacks impacting biomanufacturing processes, learning from this experience is of paramount importance. More recently, Roche and Bayer reported cyberattacks from the Winnti malware attributed to hackers in China, but were both able to detect the attack before any sensitive information could be stolen (Rees, 2019).

company functions (e.g., drug discovery and development, process design, manufacturing, quality control, clinical trials, real-world evidence). While such systems can help centralize large amounts of information, there are increasing concerns over data security and concentrating risks on a single network.

The risks and implications of cyberbiosecurity events, such as in the case of Merck & Co., may be underappreciated, especially as the role of biologic therapies across a range of conditions becomes increasingly important for meeting healthcare needs. From a manufacturing perspective, the consequences include occupational hazards, damage to equipment, batch failure leading to loss of product, and theft of IP. Regulatory burden could increase as manufacturers are required to re-establish compliance after cyberattacks, re-qualify equipment or re-validate processes. Shortages or stock-outs of medicines can lead to a loss of public trust in institutions like hospitals or the pharmaceutical industry, as well as financial burden (Caulder et al., 2015). From a patient perspective, interruptions in the supply of biologic medicines could be life threatening. The potential consequences of a cybersecurity breach range from sudden, catastrophic events such as a plant shutdown to subtler deviations in quality that introduces hard-to-detect risks into the process and increases likelihood of lot failure.

The biopharmaceutical industry is generally considered a high-value, capital-intensive and critical industry, making it an attractive target for extortionists. The batch production model for biologic therapies, vaccines and recombinant proteins, in particular, physically concentrates revenue centers since production takes place at large scales. This makes the industry vulnerable, as companies may have few runs throughout the year that each last several weeks and any form of interruption in production can damage a significant fraction of the yearly output.

While large stainless steel bioreactors have been the industry standard, there is a shift toward more flexible, single use systems that enable faster turnaround and response to uncertain demand, especially as precision medicines become available for smaller patient populations. As the industry considers more advanced manufacturing models (e.g., continuous manufacturing, real-time feedback control, etc.), close attention must be paid to the principles of information security. To identify interventions that can build resilience against potential cybersecurity threats, vulnerabilities in today's manufacturing operations, as well as future operational settings, need to be more closely examined.

DIGITAL INFORMATION FLOW IN BIOMANUFACTURING

Information exchange of highly sensitive data can be seen across the entire biomanufacturing value chain. Since a typical biomanufacturing company and the corporate network (e.g., vendors, contract manufacturing organizations) it operates in have numerous possible vulnerabilities, an important first step for these organizations is mapping risks. The following is a general, though not exhaustive, schema to help identify possible cyberattack vulnerabilities of a biomanufacturing facility. Organizations should engage experts to determine their individual security needs as they pertain to unique product types, manufacturing requirements, patient populations, regulatory jurisdictions, and geographies.

A typical biomanufacturing plant makes use of a wide range of cyber-physical systems such as sensors, actuators, programmable logic controllers (PLCs), distributed control systems (DCSs), and (in some cases) supervisory control and data acquisition (SCADA) systems (Sokolov et al., 2017). Sensors and actuators are the electronic components that take measurements of specific parameters (e.g., pH, liquid level) and execute physical responses, such as opening valves or starting or stopping a physical process. These systems are often dictated by mathematical models and algorithms with pre-determined responses based on measurements. **Figure 1** illustrates the role these systems play within standard biomanufacturing operations for the production of monoclonal antibodies.

PLCs are interfaces between specialized machinery, such as bioreactors or chromatography skids, and users. They often have specialized operating systems, sometimes with limited input interfaces, that allow them to perform dedicated functions such as integrating and displaying sensor information. They may also give feedback or automated commands that cause the system to continuously perform within preprogrammed parameters such as temperature, gas saturation, or solvent mix. PLCs have previously been overlooked in cybersecurity plans, with little awareness from manufacturers that controllers directly connected to the internet are all searchable using a single search engine, SHODAN (Wang et al., 2015). Alarmingly, SHODAN allows searchers to easily filter by machines that have retained their default security credentials. In 2011, the US Department of Homeland Security issued warnings through the Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) that nearly all PLCs are vulnerable to hackers. For instance, attackers may cause sensors to report false data or modify algorithms in control systems in ways that can jeopardize product quality, damage manufacturing equipment, and potentially induce occupational hazards.

DCSs are becoming increasingly important to the overall functioning of the production line, where multiple systems have to be coordinated to achieve the desired finished product while preventing waste or accident. They may display and integrate data from PLCs, hold plans or models, perform calculations, and allow supervisory control via input from plant workers. The DCSs allows the system or human supervisors to execute controls that affect the speed and quality of production. Because these systems can include multipurpose computers, they contain a rich amount of organizational data, and they are vulnerable to the wide array of cyberattack vectors that can affect any cyber-physical system. Some specialized hardware components for these systems could take months to replace if damaged, while re-qualifying equipment or re-validating processes can lead to lengthy supply disruptions.

A SCADA system is used in complex or distributed manufacturing systems, where a central command center issues controls and receives feedback from remote sites where physical manufacturing processes are taking place. This type of network is used in biomanufacturing for large, complex operations, or in cases where manufacturing needs to take place close to the point of care. These networks not only integrate information from the plant itself, but may also tie in to supply chain logistics or transmit information over large distances on the internet.

NEXT GENERATION MANUFACTURING

The biomanufacturing landscape is rapidly changing, partly due to technological advancements leading to process intensification, miniaturization, and automation, as well as to increased digitization of process controls more generally.

Continuous Manufacturing

Continuous manufacturing allows for a fully automated end-to-end assembly line from raw materials to products, compared to traditional batch manufacturing that requires intervention between steps of the process. Under this manufacturing paradigm, raw materials are fed into a process train and finished products removed from the other end in a continuous manner. These allow for more control over process parameters and can be run 24/7 to reduce production time. Continuous manufacturing may present additional benefits such as reducing the likelihood of costly batch dumping and real-time release of final product (National Academies of Sciences Engineering Medicine, 2019). Nevertheless, these systems present new challenges with regards to regulatory compliance and increased reliance on sensor technology for analysis of critical quality attributes. The uptake of continuous manufacturing is leading to smaller-footprint facilities with lower capital costs and thus further promote decentralization of production.

Distributed Manufacturing

Historically, the biopharmaceutical industry has concentrated manufacturing of biologics to one or few geographic locations to take advantage of economies of scale and make up for the large capital investment required in stainless-steel plants. More recently, there has been growing interest toward dividing production across multiple sites or geographic regions. A shift toward distributed manufacturing has partially been driven by the need for production systems that are more responsive to changing demand and patient-specific needs. More distributed systems, enabled by single-use technology and other advances in biomanufacturing, will make facilities more flexible and modular. Such systems rely increasingly on automation and digital networks to ensure replicability of manufacturing quality across sites, while reducing delivery time of products. While offering these potential benefits, they also introduce complex organizational and regulatory challenges, especially with regards to the cybersecurity of increasingly connected digital systems (Harrison et al., 2018).

CONSIDERATIONS FOR EMERGING BIOLOGIC PRODUCTS

ATMPs are unique in that they can require a high level of personalization and customization that could make large-scale manufacturing impractical. For these therapies, cells can either be harvested from a patient, modified, and returned to the patient (autologous) or cells originating from a single donor provide treatments to large numbers of patients (allogeneic). While the recent approvals of chimeric antigen receptor-modified T cells

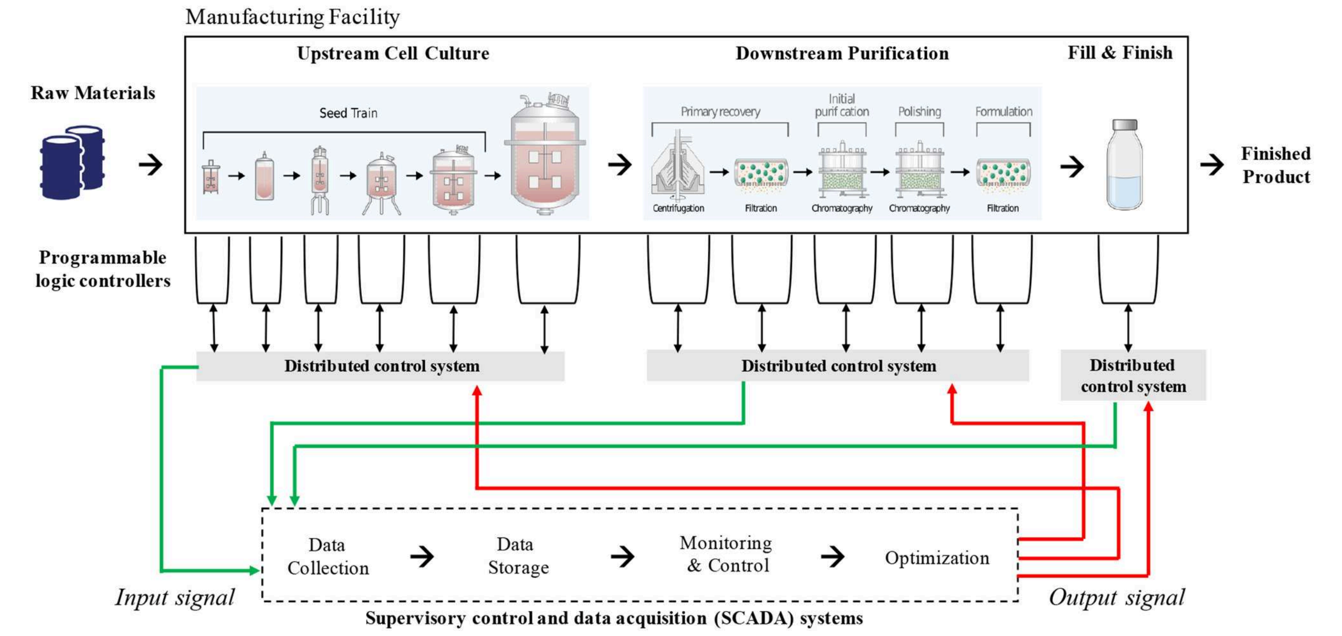


FIGURE 1 | Information flow in typical biomanufacturing operations for the production of monoclonal antibodies. The schematic indicates various cyber-physical systems that interact with each other to maintain process control. The data generated and processed by the SCADA is largely confined within a manufacturing facility.

(CAR T) spur investment into the development and production of ATMPs for a variety of new indications, manufacturing processes are not currently optimal and will likely evolve in coming years. Unique manufacturing challenges arise due to patient-specific requirements, input material variability, process-related features, and short shelf-life, amongst other factors. The production of autologous cell therapies, for example, in both a centralized and distributed manufacturing model makes use of a more complex digital information flow than that presented in **Figure 1**.

The network of facilities involved in the production and distribution of biologic therapies, as well as flow of information (data, raw materials and finished products) between them is shown in **Figure 2**, which maps both traditional biopharmaceutical and advanced (e.g., ATMP) manufacturing. Production of increasingly personalized therapies using advanced manufacturing models leads to more complex exchange of information and materials that may make these activities more susceptible to interruptions. Clinics serve as the starting point for collecting cells through apheresis and endpoint for infusion of the final product. In between, patient-specific input materials are transported to a centralized or separate manufacturing units, each with a complex set of cyber-physical systems that maintain process control. Information exchange across the network (e.g., patient, clinic, manufacturing site(s), supply chain) demonstrates the added physical and digital complexity for manufacturing these emerging therapies, thus increasing vulnerability to cyberattacks.

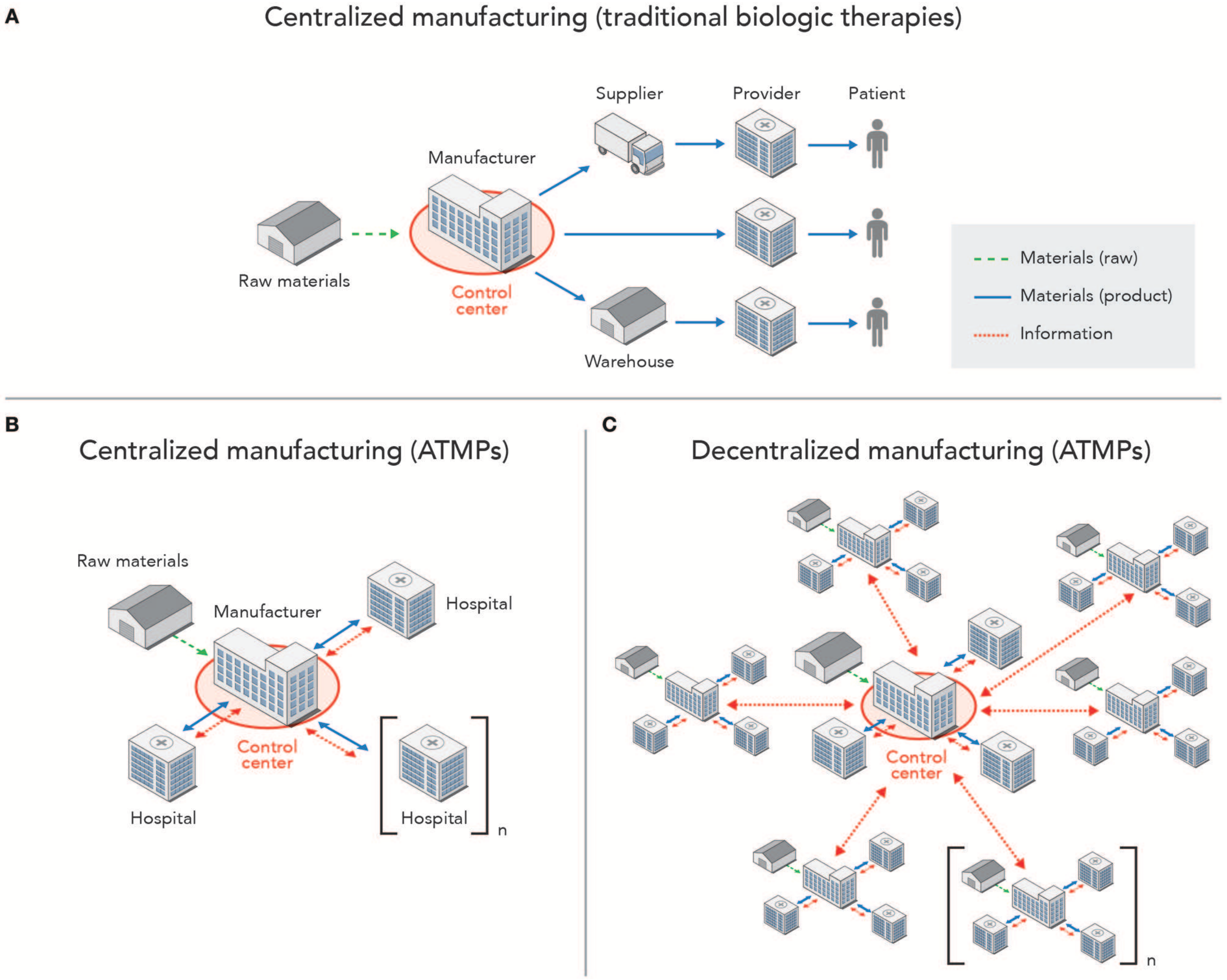


FIGURE 2 | High-level representation of information flow, including raw materials and finished products, across a network of manufacturing sites, patient providers, and control centers for monitoring of SCADA systems. **(A)** Centralized manufacturing of traditional biologic products, such as monoclonal antibodies as indicated in **Figure 1**, with a unidirectional flow of materials from the manufacturing facility to providers (e.g., pharmacies, hospitals, retailers), while information on the product is tracked across the supply chain, and a control center is housed within the facility. **(B)** Centralized manufacturing of ATMPs, which requires information and material exchange with each patient found in one or more hospitals, while the control center is still housed within the manufacturing facility. **(C)** Decentralized production of ATMPs with multiple manufacturing sites that each interface with a control center (located within one of the manufacturing facilities in the network or standalone) that monitors and manages the SCADA systems found in each facility.

Cyber-physical systems that automate manufacturing steps will play an important role in ensuring well-controlled, consistent processes while maintaining high drug quality through centralized control centers that aggregate and analyze data to inform decisions. These therapies involve complex logistics for the collection and delivery of cells to and from patients, with tight turnaround times and coordination of activities at the clinical and manufacturing sites. The need for batch release close to real-time will likely lead to additional automated procedures for validation of product quality. Data continues to be important for assurance of product quality, but there is need to better integrate pre-process, in-process and release data to execute controls across all sites and not just the one where an observation arises (Harrison et al., 2017).

Increasing use of digital systems for ATMPs, whether to monitor product quality or manage data across the product value chain, brings with it the risk of further exposing manufacturing systems to cyberattacks. These manufacturing networks feature geographically-distributed signal input and output, multiple distributed human-machine interfaces, and often, explicit tracking and use of patient data. The increasing amount and type of transmitted data opens up opportunities for malicious attackers to steal sensitive information such as patient or process information and extort money through ransomware. Additionally, high levels of variability expected in data coming from numerous distributed production systems may make it difficult to detect more subtle risks and intrusions from cyberattacks.

ENSURING A RESILIENT BIOMANUFACTURING INDUSTRY

The biopharmaceutical manufacturing industry continues to be vulnerable to cyberattacks because of the prevailing misconception that cybersecurity concerns can be dealt with using IT solutions alone and from incomplete awareness of the type and level of perceived risks, as well as limited time and resources (Kalyvas et al., 2017). Additionally, small startups in the industry may be especially susceptible since they typically run with the leanest possible staffing and resources to address cybersecurity might be limited. However, as the Merck & Co. incident shows, highly connected industries can become collateral damage as worms travel indiscriminately across systems, so each company is only as secure as their most vulnerable partner.

What are some steps that the biopharmaceutical industry should consider? In response to the increased threat and economic impact of cyberattacks, in 2014 the National Institute of Standards and Technology (NIST) released a framework for improving the national cybersecurity infrastructure. Meant to address a broad range of cybersecurity risks and applicable to organizations of all sizes and all kinds, the framework structures its recommendations into a five-step plan: identify, protect, detect, respond, and recover (National Institute of Standards and Technology, 2014). Applying this framework to the unique risks faced by biomanufacturers, firms should first identify and map their potential attack surface, from corporate workstations to PLCs which can impact normal bioprocess operations. They should institute protections on all of these surfaces, such as implementing firewalls, changing default security credentials, encrypting sensitive information, and implementing available security features. To detect incidents in a timely fashion, organizations should implement intrusion detection systems and monitoring protocols. Organizations should also have emergency response plans in place with clear lines of command and reporting. Finally, they should give thought to their recovery strategy, including mapping where offline backups of critical data and system states are stored.

As advanced manufacturing systems are increasingly considered, both in response to cost pressures and due to the unique requirements of emerging therapeutic modalities, adopting and scaling a comprehensive cybersecurity plan will be a challenge. Attack surfaces are larger and exist in different forms across the information value chain, from process data interfaces to clinical data systems. With more units digitally connected, entry points can make the entire system vulnerable to attack. The tradeoffs that emerge when considering advanced manufacturing options (e.g., greater exposure to cyber threats vs. operational gains) indicate that next-generation manufacturing may be appropriate for some but not all applications and influenced by factors beyond manufacturing (e.g., corporate culture, geography). Therefore, special attention is needed to explore the unique changes ongoing in the biomanufacturing industry and the implications they will have on ensuring manufacturing security.

CONCLUSION

Understanding the full spectrum of cybersecurity risks, including their relative likelihood and impact, across cyber-physical systems employed in biomanufacturing continues to be a challenge. This knowledge is important to proactively implement measures that will mitigate the risk and impact of cyberattacks. This requires a systematic approach to securitization, forward-looking and adaptive planning to best prepare for current and future risks, as well as promoting an industry-wide culture to address risks before they become emergencies. Suggestions have been made for greater investments in training employees, shifting the culture from one of loose self-regulation to heightened attention, and for industry to work more closely with regulators to design and implement safeguarding policies (Peccoud et al., 2018). With increasing use of complex models for advanced manufacturing, academia can play an important role in developing design principles and tools that can safeguard against cyberattacks.

Across the biomanufacturing industry, cyberattacks are experienced differently and few have been reported. Nevertheless, there are shared experiences and lessons learned that can make the entire industry safer and more resilient to a plethora of cybersecurity threats. Encouraging pre-competitive, multi-stakeholder collaboration on the best ways to prevent and detect multidimensional risks can promote knowledge sharing and improved security systems across the entire industry in ways that safeguard business interests and patient well-being. Since 2011, the Consortium on Adventitious Agent Contamination in Biomanufacturing (CAACB), a biopharmaceutical industry consortium housed at the Massachusetts Institute of Technology's Center for Biomedical Innovation, has worked to confidentially collect and anonymize data on virus contaminations in cell culture operations from Consortium-member companies. A similar approach could be taken to better understand and learn from cyberbiosecurity events across industry to move toward advanced manufacturing models in a united and safe way.

As the industry increasingly considers advanced manufacturing, especially for new therapeutic modalities, cyberbiosecurity needs to take a central role in in the design of digital strategies, business models, technologies, standards and regulations that ensure supply security. Emerging trends toward more continuous, single-use, and decentralized manufacturing will have unique implications, including unintended consequences, that will reshape the cyberbiosecurity landscape. Working together to build foresight on future potential risks will be key to turning uncertainties into opportunities in ways that safeguard biomanufacturing operations and improve access to care.

AUTHOR CONTRIBUTIONS

DG, SS, JW, and SLS all contributed equally to the design, research, writing, and review involved in the development of this manuscript.

REFERENCES

Babiceanu, R. F., and Seker, R. (2016). Big data and virtualization for manufacturing cyber-physical systems: a survey of the current status and future outlook. *Comp. Industry*. 81, 128–137. doi: 10.1016/j.compind.2016.02.004

Barrett, D. (2015). *U.S. Plans to Use Spy Law to Battle Corporate Espionage*. The Wall Street Journal. Available online at: https://www.wsj.com/articles/u-s-plans-to-use-spy-law-to-battle-corporate-espionage-1437688169 (accessed August 11, 2019).

Carman, A. (2014). *Dragonfly Malware was Designed to Target Pharmaceutical Companies*. SC Magazine. Available online at: https://www.scmagazine.com/home/security-news/dragonfly-malware-was-designed-to-target-pharmaceutical-companies/ (accessed August 11, 2019).

Caulder, C. R., Mehta, B., Bookstaver, P. B., Sims, L. D., and Stevenson, B. (2015). Impact of drug shortages on health system pharmacies in the southeastern United States. *Hosp. Pharm.* 50, 279–286. doi: 10.1310/hpj5004-279

Erman, M., and Finkle, J. (2017). *Merck Says Cyber Attack Halted Production, Will Hurt Profits*. Reuters. Available online at: https://www.reuters.com/article/us-merck-co-results/merck-says-cyber-attack-halted-production-will-hurt-profits-idUSKBN1AD1AO (accessed August 11, 2019).

Goodin, D. (2017). *Tuesday's Massive Ransomware Outbreak was, In fact, Something Much Worse*. Ars Technica. Available online at: https://arstechnica.com/information-technology/2017/06/petya-outbreak-was-a-chaos-sowing-wiper-not-profit-seeking-ransomware/ (accessed August 11, 2019).

Harrison, R. P., Rafiq, Q. A., and Medcalf, N. (2018). Centralised versus decentralised manufacturing and the delivery of healthcare products: a United Kingdom exemplar. *Cytotherapy*. 20, 837–890. doi: 10.1016/j.jcyt.2018.05.003

Harrison, R. P., Ruck, S., Medcalf, N., and Rafiq, Q. A. (2017). Decentralized manufacturing of cell and gene therapies: Overcoming challenges and identifying opportunities. *Cytotherapy*. 19, 1140–1151. doi: 10.1016/j.jcyt.2017.07.005

Helu, M., and Hedberg, T. (2015). "Enabling smart manufacturing research and development using a product lifecycle test bed," in *Procedia Manufacturing* (Gaithersburg, MD).

Kalyvas, J. R., Ridley, E. R., Overly, M. R., Howell, C. T., Rathburn, J. L., Millendorf, S. M., et al. (2017). *Cybersecurity in the Pharma, Biotech, and Medical Devices Industries: Protecting Your IP and Confidential Information in Cyberspace*. Foley & Lardner LLP. Available online at: https://www.foley.com/en/insights/publications/2017/03/cybersecurity-in-the-pharma-biotech-and-medical-de (accessed August 11, 2019).

Marsh, S. (2018). *US Joins UK in Blaming Russia for NotPetya Cyber-Attack*. The Guardian. Available online at: https://www.theguardian.com/technology/2018/feb/15/uk-blames-russia-notpetya-cyber-attack-ukraine (accessed August 11, 2019).

Morag, N. (2014). *Cybercrime, Cyberespionage, and Cybersabotage: Understanding Emerging Threats. Colorado Technical University: College of Security Studies*. Available online at: https://www.coloradotech.edu/media/default/CTU/documents/resources/cybercrime-white-paper.pdf (accessed August 11, 2019).

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J., (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

National Academies of Sciences Engineering and Medicine (2015). *Safeguarding the Bioeconomy: Applications and Implications of Emerging Science*. Available online at: https://www.ehidc.org/sites/default/files/resources/files/Safeguarding%20the%20Bioeconomy_II_Recap%20Final%20090815.pdf (accessed August 11, 2019).

National Academies of Sciences Engineering and Medicine (2019). "Continuous manufacturing for the modernization of pharmaceutical production," in *Proceedings of a Workshop* (Washington, DC).

National Audit Office (2018). *Investigation: WannaCry Cyber Attack and the NHS*. https://www.nao.org.uk/wp-content/uploads/2017/10/Investigation-WannaCry-cyber-attack-and-the-NHS.pdf (accessed August 11, 2019).

National Institute of Standards and Technology (2014). *Framework for Improving Critical Infrastructure Cybersecurity*. Available online at: https://www.nist.gov/sites/default/files/documents/cyberframework/cybersecurity-framework-021214.pdf (accessed August 11, 2019).

Osborne, C. (2018). *US Hospital Pays $55,000 to Hackers After Ransomware Attack*. ZDNet. Available online at: https://www.zdnet.com/article/us-hospital-pays-55000-to-ransomware-operators/ (accessed August 11, 2019).

Panda Security. (2017). *NotPetya Returns as "Bad Rabbit." Panda Security*. Available online at: https://www.pandasecurity.com/mediacenter/pandalabs/notpetya-bad-rabbit/ (accessed August 11, 2019).

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2018). Cyberbiosecurity: from naive trust to risk awareness. *Trends Biotechnol.* 36, 4-7. doi: 10.1016/j.tibtech.2017.10.012

Rees, V. (2019). *Roche Confirms Cyber-Attack from Winnti Malware. Europeon Pharmaceutical Review*. https://www.europeanpharmaceuticalreview.com/news/95107/roche-confirms-cyber-attack-from-winnti-malware/ (accessed August 11, 2019).

Research America. (2016). *U.S. Investments in Medical and Health Research and Development 2013-2015*. Available on;ine at: https://www.researchamerica.org/sites/default/files/2016US_Invest_R%26D_report.pdf (accessed August 11, 2019).

Sackner-Bernstein, J. (2017). Design of hack-resistant diabetes devices and disclosure of their cyber safety. *J. Diab. Sci. Technol.* 11, 198–202. doi: 10.1177/1932296816678264

Sokolov, M., Feidl, F., Morbidelli, M., and Butté, A. (2017). "Future challenges in BioPharma: the role of big data and digitalization technologies for drug manufacturing," in *Presentation Presented at PharmaTalk* (Berlin). Available online at: https://www.iottalk.eu/wp-content/uploads/2017/06/07.-Alessandro-Butte-ETH-Zurich.pdf

Symantec Security Response. (2014). *Dragonfly: Western Energy Companies Under Sabotage Threat*. Symantec. Available online at: https://www.symantec.com/connect/blogs/dragonfly-western-energy-companies-under-sabotage-threat

Symantec. (2018). *New Orangeworm Attack Group Targets the Healthcare Sector in the U.S., Europe, and Asia. Cyber Espionage Targets Sensitive Data*. Symantec. Available online at: https://www.symantec.com/blogs/threat-intelligence/orangeworm-targets-healthcare-us-europe-asia

U S Department of Health and Human Services Office for CivilRights (2019). *Breach Portal: Notice to the Secretary of HHS Breach of Unsecured Protected Health Information*. Available online at: https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf (accessed August 11, 2019).

United States Securities and Exchange Commission (2018). *Form 10-K: Commission File No. 1-6571*. Merck & Co. Inc. Available online at: https://www.sec.gov/Archives/edgar/data/310158/000031015818000005/mrk1231201710k.htm (accessed August 11, 2019).

U. S. Federal Drug Administration (2018). *Statement from FDA Commissioner Scott Gottlieb, M.D. on FDA's Efforts to Strengthen the Agency's Medical Device Cybersecurity Program as Part of its Mission to Protect Patients*. FDA. Available online at: https://www.fda.gov/news-events/press-announcements/statement-fda-commissioner-scott-gottlieb-md-fdas-efforts-strengthen-agencys-medical-device (accessed August 11, 2019).

Wang, L., Törngren, M., and Onori, M. (2015). Current status and advancement of cyber-physical systems in manufacturing. *J. Manufac. Syst*. 37, 517–527. doi: 10.1016/j.jmsy.2015.04.008

Zhong, R. Y., Xu, X., Klotz, E., and Newman, S. T. (2017). Intelligent manufacturing in the context of industry 4.0: A review. *Engineering* 3, 616–630. doi: 10.1016/J.ENG.2017.05.015

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KL declared a past co-authorship with one of the authors SLS to the handling Editor.

POLICY AND PRACTICE REVIEWS
published: 24 April 2019
doi: 10.3389/fbioe.2019.00086

Next Steps for Access to Safe, Secure DNA Synthesis

James Diggans and Emily Leproust*

Twist Bioscience Corporation, San Francisco, CA, United States

The DNA synthesis industry has, since the invention of gene-length synthesis, worked proactively to ensure synthesis is carried out securely and safely. Informed by guidance from the U.S. government, several of these companies have collaborated over the last decade to produce a set of best practices for customer and sequence screening prior to manufacture. Taken together, these practices ensure that synthetic DNA is used to advance research that is designed and intended for public benefit. With increasing scale in the industry and expanding capability in the synthetic biology toolset, it is worth revisiting current practices to evaluate additional measures to ensure the continued safety and wide availability of DNA synthesis. Here we encourage specific steps, in part derived from successes in the cybersecurity community, that can ensure synthesis screening systems stay well ahead of emerging challenges, to continue to enable responsible research advances. Gene synthesis companies, science and technology funders, policymakers, and the scientific community as a whole have a shared duty to continue to minimize risk and maximize the safety and security of DNA synthesis to further power world-changing developments in advanced biological manufacturing, agriculture, drug development, healthcare, and energy.

Keywords: biosecurity, synthetic biology, DNA, cyberbiosecurity, policy

OPEN ACCESS

Edited by:
Stephen Allen Morse,
Centers for Disease Control and Prevention (CDC), United States

Reviewed by:
Gregory D. Koblentz,
George Mason University, United States
Nancy D. Connell,
Bloomberg School of Public Health, Johns Hopkins University, United States

****Correspondence:***
James Diggans
jdiggans@twistbioscience.com

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *21 January 2019*
Accepted: *08 April 2019*
Published: *24 April 2019*

Citation:
Diggans J and Leproust E (2019) Next Steps for Access to Safe, Secure DNA Synthesis.
Front. Bioeng. Biotechnol. 7:86.
doi: 10.3389/fbioe.2019.00086

INTRODUCTION

In 2010, the United States Department of Health and Human Services (HHS) published the Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA (U. S. Department of Health Human Services, 2010). The Guidance provided a set of recommended practices to companies synthesizing double-stranded DNA to encourage such companies to screen both their customers and requested sequences. Several of the largest DNA synthesis companies came together to form the International Gene Synthesis Consortium (IGSC), a trade industry organization intended to promote the beneficial application of gene synthesis technology while safeguarding biosecurity.

The IGSC published the Harmonized Screening Protocol (International Gene Synthesis Consortium, 2009) to provide additional tactical detail around the implementation of Guidance-compliant customer and sequence screening. The Protocol specifies that synthetic gene sequence orders will be screened against the IGSC's Regulated Pathogen Database (RPD), a data set assembled and maintained by the IGSC of sequences and organisms subject to regulatory control or licensing. The Protocol further specifies that IGSC companies will only supply genes from regulated pathogens to "*bona fide* government laboratories, universities, non-profit research institutions, or industrial laboratories demonstrably engaged in legitimate research." Since its initial publication, the Protocol has been updated only once (International Gene Synthesis Consortium., 2017) to (among other minor edits) add language affirming that IGSC member companies agree not to synthesize any sequence with "best match" to variola, the virus that causes smallpox, as the disease was declared eradicated by the WHO in 1980. In addition to the Protocol, the IGSC has also developed an extensive onboarding process for potential new members to assist companies and institutions as they build new screening systems.

In the years since the publication of the Guidance, both the DNA synthesis industry and the larger synthetic biology community have rapidly advanced in terms of capability and scale. These advances create new opportunities to revolutionize many industries—from healthcare to industrial chemicals and even digital data storage. With new capabilities come new challenges to the recommendations originally spelled out in the Guidance. As the trajectory of technological advancement will inevitably continue to steepen, here we visit potential options for next steps to advance and continue to secure the manufacture of synthetic DNA and prevent the risk of misuse.

Twist Bioscience (a member company and officer of the IGSC) has witnessed first-hand how challenging some of the Guidance recommendations can become at increasing scale. Those difficulties must be surmounted while maintaining customer and sequence screening accuracy and still achieving the tight delivery timelines demanded by fierce competition within the global DNA synthesis industry.

As scale drives down cost per base pair, the relatively fixed cost of screening plays a more direct role in overall price. These costs are driven by both customer and sequence screening—commercially-available customer screening solutions still require a great deal of manual review of false positive findings. These false positives create a floor on the possible reduction in labor cost of new customer onboarding. Current sequence screening algorithms are computationally expensive and, given the high false positive rate, the results of sequence screening can be complicated to interpret. These generally require a PhD in bioinformatics both for implementation as well as day to day interpretation of hits. This makes scaling interpretation, in the absence of high-quality sequence annotation, a very expensive proposition.

Evolving technologies have blurred the lines between the gene- and oligo-length synthesis products originally addressed in the Guidance. These include ever-simpler methods for the assembly of pools of oligo-length DNA into gene-length DNA and the use of truly massive oligo pools for data storage. The data storage use case, in particular, will drive a substantial global increase in the number of unique oligo sequences under manufacture, making it ever easier to acquire the oligo-length sequences necessary to assemble genes that would otherwise be subject to regulatory control.

EVOLVING INDUSTRY BEST PRACTICES

We believe continued forward-thinking improvements in the biosecurity safety net provided by DNA synthesis order screening will require participation from all interested parties: synthesis companies themselves, policy makers, science and technology funders (both public and private), and the broader synthetic biology community.

Gene-Length Sequence Screening Performance

The Guidance and the IGSC have together accomplished a great deal in harmonizing the screening practices of the largest synthesis companies. The current IGSC onboarding protocol for new members even includes a set of test sequences to ensure that prospective member institutions have built their custom sequence screening systems with a solid level of accuracy. It is challenging, however, to determine when a custom-built screening system is "good enough"—especially given that the details of each screening implementation remain private to the implementing company. In addition, the recommendations in the Guidance do not specify particular performance metrics in terms of overall sensitivity and specificity or the degree to which sequence alteration or the source of annotation should impact screening results.

This is not the fault of the Guidance—it is extremely difficult to express in the abstract a set of performance characteristics for a system intended to screen the universe of all possible sequences. The cybersecurity and defense communities, facing similar challenges of performance estimation for complex systems, have turned to *red teaming* as a way of answering whether a given system is sufficient to accomplish a protective goal (Zhang and Gronvall, 2018). The best way to estimate whether a skilled adversary can bypass a system is to ask skilled individuals to attempt to do just that. Previous recommendations (Koblentz, 2017) have explicitly called for IGSC companies to regularly test procedures or submit to third-party audits; we believe regular red teaming by a sophisticated third party is an effective means to address these concerns. Twist has recently engaged in an extensive red teaming of our sequence screening system (publication in review) and shared the results with other IGSC members to help further improve our respective systems. We strongly recommend that synthesis companies engage in periodic red teaming as a means of assessing evolving risk of vulnerabilities in screening systems.

Red teaming has additional secondary value: sequences shown to bypass a screening system then serve as effective regression tests during follow-on software development once vulnerabilities have been patched. Regression testing is a software testing paradigm (Yoo and Harman, 2012) designed to ensure that future changes to software systems do not create new ways for previously discovered vulnerabilities to be exploited. Building and scaling a modern sequence screening system is a complex undertaking and requires using distributed computing and third-party annotation resources, both of which increase the risk of regressions during software development and maintenance. Consistent regression testing along with a suite of edge-case test sequences can help manage this risk.

Screening Oligo-Length Sequences

The 2010 Guidance set a lower bound of 200 nucleotides on the length of sequence with "best match" to organisms appearing on any of the various regulatory control lists. This was intended to strike a balance between ensuring safe manufacture of gene-length sequences while also avoiding the burden of screening for manufacturers of shorter DNA sequences. In the intervening years, however, capacity for generating enormous, diverse pools of oligo-length sequences has grown (Organick et al., 2017) while lower-cost methods for assembling high-quality, gene-length sequences from oligo pools have been developed and matured (Plesa et al., 2018). Together, these two factors create a potential vulnerability: what would be considered controlled for gene-length synthesis under current regulatory and technical systems would be permitted for synthesis as an oligo pool and could be converted into a gene length sequence by assembly in a modestly equipped molecular biology laboratory.

Proposals for screening shorter DNA sequences have been accompanied in the past by a fear of high false positive rates. This would be true were individual oligos subject to screening one at a time—we propose instead that collections of individual oligo orders and oligo pools first be subject to computational *de novo* assembly (i.e., *in silico* assembly). Such techniques (Bonham-Carter et al., 2014; Nimmy and Kamal, 2015) from the Next Generation Sequencing (NGS) community allow for computationally efficient answers to the question of actual interest: *what could I assemble (in vitro) out of this pool of short sequences*? The output from *de novo* assembly methods are longer "contig" (i.e., contiguous) sequences. These contigs should then be subjected to standard gene-length sequence screening; any red flag alert for a contig should trigger customer follow-up identical to that in the Guidance for gene-length sequences.

RESEARCH FUNDING PRIORITIES

Research funding by governments and other institutions can play a powerful role in making customer and sequence screening easier to build or acquire and more efficient (and therefore less costly to operate) while increasing the accuracy of risk estimation.

Predicting Risk in Context

The Guidance and all current sequence screening implementations focus on determining whether a given sequence is a "best match" to an entry on a list of organisms subject to regulatory control. These lists include the U.S. Federal Select Agent Program (FSAP) and the Australia Group treaty for harmonized export control. Such lists of organisms, in the context of sequence screening, are generally proxies for a broader goal: determining whether a given ordered sequence could be used to cause significant harm.

For a regulatory control regime to focus on this much more salient challenge, we must move beyond lists of known pathogens and instead focus on the biological context and known "routes to harm." These can be as simple a single protein (e.g., in the case of ricin) or as complex as the potentially hundreds of genes required for a bacterial pathogen (e.g., the genes required by *Francisella tularensis* to cause tularemia). This annotation requires a committed, ongoing effort to catalog, in detail, the ways in which proteins and genetic networks can be used to cause harm in contexts subject to regulatory control. The knowledge of these mechanisms and the genes they require is highly specialized and diffuse across academic, government, and industrial experts. We understand the assembly of this knowledge in a single, shared location to be both incredibly important and incredibly challenging.

Sustained funding and commitment will be required to build and maintain a database of risk-associated sequences, their known mechanisms of pathogenicity and the biological contexts in which these mechanisms can cause harm. This database (or at a minimum a screening capability making use of this database), to have maximum impact on global DNA synthesis screening, must be available to both domestic and international providers. Arguments have previously been made that such a collection would make misuse of biology easier for bad actors. Modern deep learning methods, while powerfully predictive, often require enormous amounts of high-quality, curated training and specialized statistical expertise to make accurate predictions on complex outcomes. Allowing access only to synthesis companies or others with a "need to know" establishes a threshold for who can work on these challenges and limits the degree of global creativity that can be applied to the challenge of predicting biological outcomes from collections of primary sequence. We believe the value provided by the collection and public dissemination of this information, in terms of empowering machine learning and other risk estimation efforts, far outweighs any increased potential for attempted misuse.

We have excellent examples of this approach in the cybersecurity community: Common Vulnerabilities and Exposures (CVE) (MITRE, 1999) and the National Vulnerability Database (NVD) (National Institute of Standards and Technology, 2000). CVE and NVD publicly catalog known vulnerabilities and code exploiting those vulnerabilities. These data are used to build ever-more-capable intrusion detection systems and to inform software development practices to avoid creation of new vulnerabilities. We believe this same paradigm would work well in a biological context.

As this database grows, additional investment in statistical methods for risk estimation will result in approaches with increasing accuracy in predicting harm. These systems should move from predicting risk on primary DNA sequences to include predicting possible harmful outcomes from genetic circuit designs or even from engineered microbial communities. The Intelligence Advanced Research Projects Agency, IARPA, is funding early work in this area via its Functional Genomic and Computational Assessment of Threat (FunGCAT) program (IARPA, 2016). We strongly encourage funding of complementary and follow-on approaches.

The metaphorical similarity to the cybersecurity domain is not, admittedly, perfect. Patching software vulnerabilities is far easier and less expensive than "patching" biological vulnerabilities via vaccines or novel medical countermeasures. This does not mean, however, that simply enumerating the genes required for a particular "route to harm" is sufficient information to enable bad actors—a flat list of genes involved in a pathogenic outcome is not a recipe. Furthermore, there are large scale efforts underway including the DARPA Pandemic Prevention Platform (P3) program (DARPA., 2017) to enable just this sort of rapid response to novel pathogens. We maintain that the upside of providing this level of detail—low-cost, uniformly accurate, peer-reviewed sequence screening—more than offsets any potential for additional information hazard.

Sharing Risk Estimation Across Companies

IGSC companies have long recognized the risk of "venue shopping"—that a bad actor intent on acquiring dangerous sequences could submit an order to multiple companies in the hope of finding a company whose screening system will permit the order. The IGSC addresses some of this risk by having each company alert the other IGSC companies to any order causing significant concern.

This still leaves a potential vulnerability in terms of an individual ordering sub-threshold sequences from multiple companies and then carrying out final assembly themselves. The only way to gain a shared awareness of this kind of activity would be to devise a system for sharing assembly and alignment data across companies. Such a system, however, would need to be hosted by a trusted third party and not disclose business-sensitive information including the underlying sequences themselves, the total volume of sequence from any individual company, or any decision-making to manufacture on the part of contributing companies.

Technical solutions to this problem could include sharing only sub-sequences (referred to as "k-mers" in bioinformatics, i.e., sub-sequences of length k) as well as more exotic mathematical methods including homomorphic encryption. Homomorphic methods (that is, methods allowing for computation on data that remains encrypted throughout) would theoretically allow for alignment of sequences to a set of controlled references without disclosing the exact composition of the query sequence (Esvelt, 2018; Titus et al., 2018). In the absence of actual homomorphic alignment methods and given recent work in pseudo-alignment for RNA-Seq data (Bray et al., 2016), we believe pseudo-alignment approaches show the most near-term promise. They operate on k-mers (rather than requiring full sequences) and scale efficiently by, paradoxically, *not* focusing on determining detailed homology-based matches of a query sequence to a database of possible origin sequences. Instead, they estimate only the likelihood that a given sequence came from a given origin sequence—the statistical "best match." This aligns precisely with the challenge posed to synthesis sequence screening.

Democratizing Access to Sequence Screening

Maximizing the security of global DNA synthesis will require an ever-larger tent as new synthesis companies are created and grow around the world to serve local or other niche markets. Building a screening system, however, can be expensive and non-trivial. Especially for companies whose business model focuses on thin margins or low volume, the current economics (even with extensive IGSC advice and support) strongly dis-incentivize screening. To lower this barrier to entry for screening, we must solve two problems: software for carrying out the screen and access to high-quality, up-to-date annotation on controlled toxins, viruses, and bacteria. The previous recommendation for ongoing commitment and public availability of a database of "routes to harm" satisfies the second of these criteria. The first could be satisfied by the creation of a small but competitive market for software-as-a-service-based solutions or even open source software allowing a company to quickly install and screen (at low volume) with high accuracy.

Open source would also allow peer-review of algorithmic approaches to screening, further insuring against the risk of software vulnerabilities driving unintended access to sequences. Open software development, however, would require access to curated screening data both to be used by the tool operationally as well as to rigorously test the implementations to ensure they cannot be subverted via clever construct design. This need for validation could create communities of individuals attempting to build sequences that might expose vulnerabilities—this, again, leverages a useful pattern in the cybersecurity world of "bug bounty" programs meant to encourage the constructive application of creativity to identify and report software weaknesses.

INTERNATIONAL NORMS AND THE SECURITY MINDSET

Long-standing efforts within the synthetic biology community raising awareness of the potential security applications of these technologies has paid dividends and should be expanded. The community must ensure that DNA synthesis companies are not seen as the only stopgap to misuse. Companies designing genetic circuits and novel organisms often are, and should continue to be, active participants in security-related threat evaluation and estimation of potential misuse of the technologies they invent, mature, and sell. We recommend that the focus of the 2010 Guidance on "know your customer" should apply more broadly and explicitly to the entire synthetic biology industry and supply chain.

In addition to building this awareness within companies, it is crucial to continue and expand education efforts on the importance of biosecurity and development of a security mindset in synthetic biology. The International Genetically Engineered Machine (iGEM) competition Safety and Security and Human Practices efforts have educated thousands of young scientists on the importance these kinds of security considerations in synthetic biology. The Engineering Biology Research Council (EBRC) in the United States recently held a Department of Homeland Security (DHS)-funded workshop focused on further improving consideration of security in synthetic biology, recommending that graduate-level scientific education should explicitly teach security awareness to young researchers.

The workshop also highlighted the potential value of asking in grant applications that, in addition to considering the *safety* implications of proposed work (i.e., how might this work accidentally harm yourself or others), applicants also demonstrate that they have thought through the *security* implications (i.e., how might this work be used to *intentionally* harm others). This can improve early awareness of broader security implications of new technologies and foster community discussion and interaction on the risk and benefit trade off and evaluation by a broader community of ethicists or other relevant experts.

Internationally, the Nuclear Threat Initiative has recently launched their Global Biosecurity Innovation and Risk Reduction Initiative intended to "develop, publicize and promote concrete and normative actions to reduce global catastrophic biological risks associated with advancements in technology" (Nuclear Threat Initiative., 2018). Such large-scale, well-funded international activities are extremely valuable in establishing and harmonizing expectations of security considerations and behavioral norms across national borders.

CONCLUSION

With increasing scale and complexity in manufacture of synthetic DNA, and in synthetic biology more broadly, comes a responsibility to ensure these technologies continue to be used responsibly. Here, we have outlined a multi-faceted approach to advance the technology, policy, educational, and social environments that help guard against potential misuse. We recommend periodic red teaming to ensure an understanding of the current performance characteristics of DNA sequence screening systems. Additional science and technology investment can build the annotation resources and algorithms necessary to continue to improve both the accuracy and affordability of screening. By lowering the cost of screening and making open source annotation resources and tools available, a much wider array of synthesis companies will be able to screen their orders.

We also recommend that the U.S. government extend guidance to include screening of oligonucleotide pools. This approach should emphasize hypothesis generation via *de novo* assembly from one or many oligo pools rather than focusing alerting on single, short sequences (which can lead to high false positive rates). We further suggest that the U.S. government guidance to "know your customer" apply broadly across the synthetic biology supply chain. In addition, we actively encourage efforts to teach and promote the evaluation of the security implications of new synthetic biology techniques or materials as part and parcel of being a practicing synthetic biologist.

Together, these steps will ensure screening and security practices scale both in terms of the rapidly growing number of global synthesis requests as well as evolving with increasing human knowledge of biological systems and functional components. This multifaceted approach will better serve our shared duty to use synthetic DNA to protect and improve the well-being of people and our planet.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was funded by Twist Bioscience Corporation.

REFERENCES

Bonham-Carter, O., Steele, J., and Bastola, D. (2014). Alignment-free genetic sequence comparisons: a review of recent approaches by word analysis. *Brief. Bioinformatics* 15, 890–905. doi: 10.1093/bib/bbt052

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527. doi: 10.1038/nbt.3519

DARPA. (2017). *Pandemic Prevention Platform (P3)*. Available online at: https://www.darpa.mil/program/pandemic-prevention-platform

Esvelt, K. M. (2018). Inoculating science against potential pandemics and information hazards. *PLoS Pathogens* 14:e1007286. doi: 10.1371/journal.ppat.1007286

IARPA (2016). *Functional Genomic and Computational Assessment of Threats (Fun GCAT)*. Available online at: https://www.iarpa.gov/index.php/research-programs/fun-gcat

International Gene Synthesis Consortium (2009). *Harmonized Screening Protocol*. Available online at: https://portal.sgidna.com/files/IGSC%20Harmonized%20Screening%20Protocol.pdf

International Gene Synthesis Consortium. (2017). *Harmonized Screening Protocol V2*. Available online at: https://genesynthesisconsortium.org/wp-content/uploads/IGSCHarmonizedProtocol11-21-17.pdf

Koblentz, G. D. (2017). The *de novo* synthesis of horsepox virus: implications for biosecurity and recommendations for preventing the reemergence of smallpox. *Health Security* 15:6. doi: 10.1089/hs.2017.0061

MITRE (1999). *CVE - Common Vulnerabilities and Exposures (CVE)*. Available online at: https://cve.mitre.org/. (accessed January 7, 2019)

National Institute of Standards and Technology (2000). *National Vulnerability Database*. Available online at: https://nvd.nist.gov/. (accessed January 7, 2019)

Nimmy, S. F., and Kamal, M. S. (2015). Next generation sequencing under *de novo* genome assembly. *Int. J. Biomath.* 08:1530001. doi: 10.1142/S1793524515300018

Nuclear Threat Initiative. (2018). *NTI Launches New Global Biosecurity Innovation and Risk Reduction Initiative*. Available online at: https://www.nti.org/newsroom/news/nti-launches-new-global-biosecurity-innovation-and-risk-reduction-initiative/. (accessed October 30, 2018).

Organick, L., Ang, S. D., Chen, Y.-J., Lopez, R., Yekhanin, S., Makarychev, K., et al. (2017). Scaling up DNA data storage and random access retrieval. *BioRxiv*:114553. doi: 10.1101/114553

Plesa, C., Sidore, A. M., Lubock, N. B., Di, Zhang, and Kosuri, S. (2018). Multiplexed gene synthesis in emulsions for exploring protein functional landscapes. *Science* 359, 343–347. doi: 10.1126/science.aao5167

Titus, A. J., Flower, A., Hagerty, P., Gamble, P., Lewis, C., Stavish, T., et al. (2018). SIG-DB: leveraging homomorphic encryption to securely interrogate privately held genomic databases.

PLoS Comput. Biol. 14:e1006454. doi: 10.1371/journal.pcbi1006454

U. S. Department of Health and Human Services (2010). *Screening Framework Guidance for Providers of Synthetic Double-Stranded DNA*. Available online at: https://www.phe.gov/Preparedness/legal/guidance/syndna/Pages/default.aspx

Yoo, S., and Harman, M. (2012). Regression testing minimization, selection and prioritization: a survey. *Softw. Test. Verification Reliabil.* 22, 67–120. doi: 10.1002/stvr.430

Zhang, L., and Gronvall, G. K. (2018). Red teaming the biological sciences for deliberate threats. *Terrorism Political Violence* 1–20. doi: 10.1080/09546553.2018.1457527

Conflict of Interest Statement: JD and EL are employed by Twist Bioscience. Twist Bioscience is a board member of the International Gene Synthesis Consortium (IGSC). The views expressed here are not necessarily those of the IGSC.

frontiers
in Bioengineering and Biotechnology

HYPOTHESIS AND THEORY
published: 07 August 2019
doi: 10.3389/fbioe.2019.00189

On DNA Signatures, Their Dual-Use Potential for GMO Counterfeiting, and a Cyber-Based Security Solution

*Siguna Mueller**

Independent Researcher, Kaernten, Austria

This study investigates the role and functionality of special nucleotide sequences ("DNA signatures") to detect the presence of an organism and to distinguish it from all others. After highlighting vulnerabilities of the prevalent DNA signature paradigm for the identification of agricultural genetically modified (GM) organisms it will be argued that these so-called signatures really are no signatures at all - when compared to the notion of traditional (handwritten) signatures and their generalizations in the modern (digital) world. It is suggested that a recent contamination event of an unauthorized GM *Bacillus subtilis* strain (Paracchini et al., 2017) in Europe could have been—or the same way could be - the consequence of exploiting gaps of prevailing DNA signatures. Moreover, a recent study (Mueller, 2019) proposes that such DNA signatures may intentionally be exploited to support the counterfeiting or even weaponization of GM organisms (GMOs). These concerns mandate a re-conceptualization of how DNA signatures need to be realized. After identifying central issues of the new vulnerabilities and overlying them with practical challenges that bio-cyber hackers would be facing, recommendations are made how DNA signatures may be enhanced. To overcome the core problem of signature transferability in bioengineered mediums, it is necessary that the identifier needs to remain secret during the entire verification process. On the other hand, however, the goal of DNA signatures is to enable public verifiability, leading to a paradoxical dilemma. It is shown that this can be addressed with ideas that underlie special cryptographic signatures, in particular those of "zero-knowledge" and "invisibility." This means more than mere signature hiding, but relies on a knowledge-based proof and differentiation of a secret (here, as assigned to specific clones) which can be realized without explicit demonstration of that secret. A re-conceptualization of these principles can be used in form of a combined (digital and physical) method to establish confidentiality and prevent un-impersonation of the manufacturer. As a result, this helps mitigate the circulation of possibly hazardous GMO counterfeits and also addresses the situation whereby attackers try to blame producers for deliberately implanting illicit adulterations hidden within authorized GMOs.

Keywords: cyberbiosecurity, DNA signatures, bio-cryptanalysis, bio-cyber hacker, insecure channel, GMO counterfeiting, cryptographic applications, knowledge-based methods

OPEN ACCESS

Edited by:
Randall Steven Murch,
Virginia Tech, United States

Reviewed by:
James Diggans,
Twist Bioscience, United States
Edward Colbert,
Hume Center for National Security and Technology, Virginia Tech, United States

****Correspondence:***
Siguna Mueller
siguna.mueller@protonmail.com

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *05 March 2019*
Accepted: *18 July 2019*
Published: *07 August 2019*

Citation:
Mueller S (2019) On DNA Signatures, Their Dual-Use Potential for GMO Counterfeiting, and a Cyber-Based Security Solution. Front. Bioeng. Biotechnol. 7:189. doi: 10.3389/fbioe.2019.00189

1. MOTIVATION

The cyber-physical nature of biotechnology raises unprecedented security concerns, and "Cyberbiosecurity" has been recognized as a critical imperative to "help safeguard the bioeconomy" (Murch et al., 2018; Peccoud et al., 2018). One of the critical new challenges concerns the gap between a (digital) description of a certain product and its actual (physical) realization. This was first illustrated by Peccoud et al. (2018) as they experienced major difficulties when trying to reproduce purported sequences of a plasmid sent in the mail. The actual expression characteristics of the plasmid were drastically different than what was expected from their description.

As noted by Peccoud et al. (2018), the security risks of the problems at the interface between the digital and biological/physical realms are profound. A related (but much more problematic) incident recently emerged in several countries of the European Union (Paracchini et al., 2017). Nowadays, many food and feed additives result from fermentation of genetically modified (GM) microorganisms. Microbial synthesis of vitamin B2 (riboflavin) often involves GM *Bacillus subtilis* production strains. According to European Guidelines (EFSA, 2011), for additives produced with GM microorganisms, it is necessary that in the final product neither the production strain nor its recombinant DNA can be detected. However, in September 2014, viable GM *B. subtilis* spores were detected in a consignment of vitamin B2 feed additives imported from China. Molecular characterization confirmed that these were not the strains that the manufacturers claimed to be using (Paracchini et al., 2017). In other words, the description of the product (as authorized within the EU) did not match the actual one (which was shown to harbor several unauthorized GM modifications).

The European Union has strict GMO regulations and testing mechanisms in place to determine unauthorized GMOs and to ensure compliance with regulations. The ones that are considered most reliable in fact offer real-time PCR detection of GMO-specific signatures (Permingeat et al., 2002; Levine, 2004; Allen et al., 2008), yet, herein, their role and functionality is challenged.

Originally, DNA signatures were invented to accurately distinguish between a target genome (or a set of genomes) and all other background genomes (Phillippy et al., 2007). For practical reasons, research has focused on balancing the tradeoff between signature sensitivity (the number of genomes that share the signature) and specificity (the number of genomes that do not possess the signature). With advancements in genetic engineering, however, it has become possible to actually insert artificial signatures (e.g., Gibson et al., 2010), whereby it has become possible to differentiate artificially modified from natural organisms.

DNA signatures based on integration sites between the transgene insert and the flanking DNA make use of this same idea. While these types of signatures have been the paradigm of GMO detection for decades, this article strongly challenges the function of such signatures, especially relative to intended manipulations.

Both traditionally and in the cyber-domain, signatures have long served as a valuable tool to guarantee the integrity and authenticity of the document being signed. However, the very concept of signatures in the cyber-realm first needed to be redressed as the Internet is susceptible to intrusions that are not existent in the traditional setting. Analogously, it is argued here, that unique signature vulnerabilities exist in the biologic domain.

A very recent study (Mueller, 2019) demonstrates that the existing DNA signature paradigm may be exploited via previously unrecognized forms of attack. It is suggested that new gene editing technologies can be used to create plants that are genetically modified in harmful ways, either in terms of their effect on the plant itself or in terms of harming those who would consume foods produced by that plant. This possibility opens up an unrecognized avenue for bioterrorism or biocrime—either by maliciously modifying a natural organism or (perhaps more perniciously) sabotaging a previously approved GMO. The role that DNA signatures play here is critical. The problem is not only that any clandestinely introduced manipulations are difficult to detect, but that the standard verification of DNA signatures leads to a false sense of security, as illegal or detrimental alterations can be made without changing the authenticating identifiers. This enables the adulterated product to pass as the original if only the identifiers are examined.

This article offers a detailed analysis of risk potentials of DNA signatures, with special focus on the identification of agricultural GMOs. Based on lessons learned from the cyber-domain, specific vulnerabilities are highlighted, and recommendations are made how these new risks can be mitigated.

1.1. Outline

Section 2 analyzes the role of DNA signatures as conceptualized in a broader framework inspired by cryptography. Section 3 describes specific risks arising in the biological realm, how conventional DNA signatures can lead to new forms of counterfeiting attacks. Section 4 considers practical issues for performing such attacks and overlays them with their cyber-based conceptualization. The combination of these two leads to specific recommendations how DNA signatures may be improved. A method how enhanced DNA can be realized through specific cryptographic tricks complemented by a suitable physical realization is described in section 5.

2. FROM CRYPTOGRAPHIC TO BIOLOGIC SIGNATURES

This section gives the necessary background how traditional signatures were recaptured in the cyber domain, to establish analogous security features. The insights derived will help distill critical vulnerabilities in the biologic domain.

2.1. What Needs to be Protected

While the cyber realm is shaping our everyday lives, its underpinnings can be traced back for many centuries. Originally, it was in the form of secret exchange of messages during times

Abbreviations: GM, genetically modified; GMO, GM organism; UGMO, unauthorized GMO; NGS, Next Generation Sequencing; WGS, Whole Genome Sequencing; ZK, Zero Knowledge; TTP, Trusted Third Party.

of war. Out of this evolved the discipline of cryptography which later branched out into various cyber related disciplines. As cryptographic insights and ideas have been an important component of cybersecurity, it is worthwhile to analyze the underlying principles, to help guide their application for DNA signatures.

2.1.1. Cryptographic Goals

According to Stinson (2005), the objective of cryptography is, "to enable two people, usually referred to as Alice and Bob, to communicate over an insecure channel in such a way that an opponent, Oscar, cannot understand what is being said." The core issue lies in the (insecure) channel, as summarized by Claude Shannon in 1948 (see MacKay, 2003), "The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point."

Although usually not conceptualized this way, it can be beneficial to rephrase essential cybersecurity principles in terms of insecure channels. Doing that will help filter out parallels as well as differences to the biologic domain. For instance, for the well-known CIA (or AIC) triad (see **Table 1**), this means:

- **Confidentiality**: The goal is to limit access to information (or, the process, and production of bioengineered products) from reaching the wrong people. Alternatively, the defining characteristics of the channel ("the message") can be seen as "who is allowed to have access." This "message" is intended to not change in the life-cycle of the entire information process.
- **Integrity**: This is about ensuring that information (or, a biomanufactured entity or process) is trustworthy, consistent, and accurate over its entire life-cycle. Alternatively, in terms of a channel, "the message" is the information (content) to be secured. This should be the same at either end of the secure channel, that is, wherever and whenever it is asked for.
- **Availability**: This feature aims at guaranteeing reliable access to information (or, any cyberbiosecurity process). Alternatively, "the message" to be secured is "what is available" (to authorized individuals). This very information should be the same across different points of the channel (wherever it is needed).

Although genetic engineering is not dealing with digital or electronic "communication" and "messages," there are critical parallels - as well as differences - that are worth investigating. Traditional (noisy/insecure) communication channels are a telephone line, a flash drive, or computer network, for example. These show that communication is not exclusively understood as "information going from one *place* to another" (MacKay, 2003). When we write a file on a flash drive, "we'll read it off in the same location - but at a later time" (MacKay, 2003). MacKay gives the example of "reproducing cells, in which the daughter cells' DNA contains information from the parent cells." This is a "noisy" channel, as this process is subject to (unintended) mutations or change.

2.2. Cryptographic Signatures

Arguably, if there has ever been a single paradigm that has been most influential in the security setting, then it has been that of digital signatures. (Written) signatures are providing a number of critical services, including non-repudiation, entity or data origin authentication, and identification. When electronic signatures were first developed, a redressing of the traditional concept was required. Consequently, they too, can now be used to guarantee analogous safety features in information-theoretic communication systems (see **Table 2**).

2.3. Information Channels at the Cyber-Biological Interface

Conceptualizing security primitives in form of insecure channels has several advantages, most notably because it directs the focus to the key players that need to be secured ("signed"). When utilizing this approach, however, **Table 3** shows unique challenges in the biologic field and demonstrates that the concept of traditional signatures may not be realizable.

It is evident that many of the biologic insecure channels are in fact "insecureable." That is, in contrast to the traditional/digital domain, it is often not plausible to assign a "fixed message" which is expected to have the same value across the channel. For instance, digitally, it is easy to depict the content of a certain communication, and securing the channel means that this same content can be obtained at both channel sides (sender and receiver). Finding an analog to this for living entries often is not possible, as such "messages" constitute living and flexible entities involving indels, SNPs, jumping genes, synergistic effects, or functional relationships between different forms of information and their environment.

2.3.1. The Problem With Signatures in "Insecureable" Channels

A critical property of signatures is to enhance an insecure channel, and to verify that it has been "secured." Traditionally, the signature on a legal document would serve as the means of this conclusion. In the opposite case, any alterations to the document would invalidate the signature, indicating that the channel has not been secured yet. The same is true for cryptographic signatures. If the message is "Send $ 5 to Account xyz," then the signature that is (mathematically) computed from this "message" would be radically different to the one obtained from the message "Send $ 5,000 to Account zyx."

In order to secure a channel it is therefore necessary to identify the underlying "message" (see **Table 3**), to sign it in its entirety, and to verify whether or not the intended "message" has been retrieved. In some circumstances, the same is doable for bioengineered entities. For instance, with an artificial plasmid, it would be possible to obtain the complete sequencing information ("the message") which may be assumed to be fixed and stable. While practically this may be costly (as it would require sequencing of the entire genome) the situation represents a "secureable" channel. Indeed, the retrieved sequencing information could be compared with an authenticated sequencing information [e.g., from an appropriate database, as suggested in (Peccoud et al., 2011, 2018; Dunlap and Pauwels, 2017)]. In this regard, the latter may even assume the role of a signature for the physical entity.

TABLE 1 | Cryptographic concepts and goals. In the cyber-domain, many of those can be addressed via digital signatures.

Crypto/cyber goals	Description	Cryptographic solution
Confidentiality	This is a service used to keep the content of information from all but those authorized to have it. Secrecy is a term synonymous with confidentiality and privacy. There are numerous approaches for providing confidentiality, ranging from physical protection (e.g., a box with a lock, a sealed envelope, or a wall-safe) to mathematical algorithms which render data unintelligible	Digital signatures, access control, hardware protection
Data integrity	This is a service which addresses the alteration of data. To ensure data integrity, one must have the ability to detect arbitrary errors, as well as manipulation by unauthorized parties. Data manipulation includes such things as insertion, deletion, and substitution	Hashing, message-authentication protocols, digital signatures
Authentication	This is a service related to identification. This function applies to both entities (e.g., a person, a credit card, an information-carrying product - including one that is biomanufactured) and information (in particular, the source of information, including its origin, date of origin, data content, time produced, etc)	Digital signatures, passwords, authentication protocols, challenge and response
Availability	Is a guarantee of reliable access (to information, computers, specific components or systems, etc) by authorized people	Updates, backups, firewalls, proxy servers, physical protection
Non-repudiation	This is a service which prevents an entity from denying previous commitments or actions. When disputes arise due to an entity denying that certain actions were taken, a means to resolve the situation is necessary	Digital signatures, public-key schemes, trapdoor functions, commitment schemes

It is suggested that these conceptions can help identify key functionalities of biologic signatures as well.

TABLE 2 | Principles and features of digital signatures as counterparts of traditional signatures (and with the intent toward their generalization to DNA signatures).

Digital signatures	
How they work	"Public-key" signatures rely on the usage of specific secrets - the keys used to generate a signature. They are generated by applying a mathematical formula or an algorithm, to scramble the information into a string of digits
Who can produce a valid signature?	Only the holder of the private (secret) key–the signer–can produce such an "electronic autograph"
Who can verify a signature?	In the public-key setting, the signature can be verified by anyone
Useful features	
They provide authenticity and enable supply chain security	For messages distributed through a non-secure channel, a properly implemented digital signature gives the receiver reason to believe the message was sent by the claimed sender
They provide data integrity and ensure anti-counterfeiting	Any change in the message after signature will invalidate that signature, which ensures the integrity of the signed data ("the message") against tampering or corrupting during transmission
They are binding	Once it is published, a signature cannot be altered or repudiated
What can be signed?	As with anything in the cyber-realm, the message is an alphanumeric string, including anything that can be represented as such (genomic information, producer information, processes used, etc)

Unfortunately, GMOs (especially in the agricultural setting) cannot always effectively be sequenced in full to obtain a fixed message that remains unchanged across time and space. Here, generating a signature could be likened to signing a blank check, or worse. **Figure 1** provides a "bio-cryptanalytical" summary of the main challenges that arise in this context.

3. THE POTENTIAL AND RISK OF LARGE-SCALE INTRUSIONS

3.1. Accidental GM Contamination and Signature Theft for Cost-Saving Purposes

The detection of the unauthorized GM *B. subtilis strain* in Europe has led to rigorous investigations to identify the unknown genetic insertions/deletions that are responsible for the significant overproduction of vitamin B2 (Barbau-Piednoir et al., 2015; Paracchini et al., 2017). The analysis by Paracchini et al. (2017) revealed genetic adulterations in form of specific indels as well as extra-chromosomal recombinant plasmids that are presumably conferring antibiotic resistance for selection purposes and stable riboflavin expression during fermentation. Correspondence with the manufacturers revealed they were relying on known GM-strains, which means that at some point some type of identification process must have been in place. The problem arose when these authentic identifiers falsely got associated with the modified strains. The same, however, could be done by bio-cyber hackers in form of intended DNA signature misuse (theft) to masquerade an unauthorized product.

3.2. DNA Signature Theft With a Malicious Intent

3.2.1. Signature Theft to Harm the Reputation of the Manufacturer

In the B2 contamination event, rigorous investigations between European and Chinese competent authorities led to the

TABLE 3 | Examples of "insecure channels" in the field of cyberbiosecurity.

Insecure channel	"The message" (what is to be secured)	Feature of a secure-able channel	Comment
DNA replication: The process of passing on a parental piece of DNA to offspring	The specific DNA sequence	The DNA sequence is the same before and after replication	Numerous cellular repair mechanism turn the potentially insecure/noisy channel into one that is secured
Artificial plasmids. These are carefully designed to lead to a specific trait. Specifics of the expressed phenotype are coded in the artificial sequences	The artificial DNA cassette	The sequence information of the artificial construct is the same, regardless of the lab or environment that it is utilized by. To be "secure-able" means that this information can be traced back to its original/intended sequence	Sequencing of the plasmid allows to reveal its complete and detailed sequencing information. While this is costly and technically demanding, this shows if the channel (the sequence information encoded by the plasmid) matches the expected sequence [as e.g., can be verified by secured databases (see Peccoud et al., 2011, 2018; Wilson et al., 2012; Dunlap and Pauwels, 2017)]
Raw data, health related information, medical databases (storage of man-made information, as opposed to sequence information in living organisms)	The digital information about medical insights, health records, etc	The digital data remain unaltered (same information regardless of when and by whom it is read out), accessible only to legitimate authorities, and whenever needed	Once the information is in place, this essentially is a cyber-problem and can therefore benefit from existing cyber-related tools
Artificial DNA sequences, DNA as information storage	The message is the information to be stored in form of artificial DNA bases	As above	Need to filter out alterations due to DNA processing. Can benefit from alignment-based methods such as distance-measures (e.g., Federhen et al., 2016) and additional coding-theory and cyber-based tools to identify, correct, or minimize any errors (see e.g., Mueller et al., 2015)
Expression of a transgene via a GMO. Targeted phenotypic trait and expression levels	(a) "The message" is the specific transgene. The channel that aims to be protected is the transgene only (b) "The message" is the entire genome. The channel that aims to be protected is the entire organism	The transgene achieves its targeted phenotypic expression, relative to its trait, expression level, and in the context of its intended (molecular, biologic, cellular) environment	(a) The phenotypic expression can be influenced by illicit genetic modifications outside of the transgene. If integrity is verified with respect to the transgene only, such covertly introduced modifications are not detected. They lie outside the specific channel (b) To obtain a secure-able channel, it needs to be the case that (1) The entire genome can be sequenced, (2) The sequencing information obtained in different contexts and circumstances always lead to the exact same sequence (possibly including predictable differences within a certain range or distance)
Modern gene-edited plants and crops (see e.g., Grohmann et al., 2019)	Unclear what the message is. This is because the intended effect is based on a range of expression levels via specific biochemical pathways, which are dependent on their context and environment (here, environment is meant across the full spectrum, from molecular to gross)	The intended outcome is a spectrum of traits, depending on the specific context and environment. Here, secure-able would mean the same spectrum of phenotypic expression, as informed by different, discrete conditions in a clearly causative way	It seems much more difficult to secure a channel like this, where there is no tangible fixed, physical message that can be identified as the key information to be protected

The key feature of insecure channels often can be formulated in terms of existing cryptographic primitives. For instance, all channels involve attributes that aim at leaving some information unaltered (integrity). Insecure channels in the cyber domain build on the salient feature that these can in fact be "secured." In the context of integrity this would mean that the original intended information can be recaptured. In cryptography, what needs to be secured is typically called "the message." It is important to note that this term has nothing to do with our contemporary usage of this word. Here, it describes the defining characteristics of the insecure channel. By identifying "the message" involving biological mediums it is found that many of the insecure channels are in fact "insecureable".

conclusion that the "the production strain must have been contaminated or switched before or during production" and that it concerned an "exceptional" and "singular" case (Paracchini et al., 2017). The genetic alterations turned the feed additive into something that is unauthorized in Europe. One can imagine that when done on purpose, such types of attacks may be performed with the explicit intent to harm the reputation of the manufacturer. Similarly, counterfeiters may try to ingress more harmful manipulations just to blame the producer. It is important to note that such attacks may not easily be detected, especially when nobody is looking for such (unknown) alterations. Nonetheless, even the advanced PCR methods

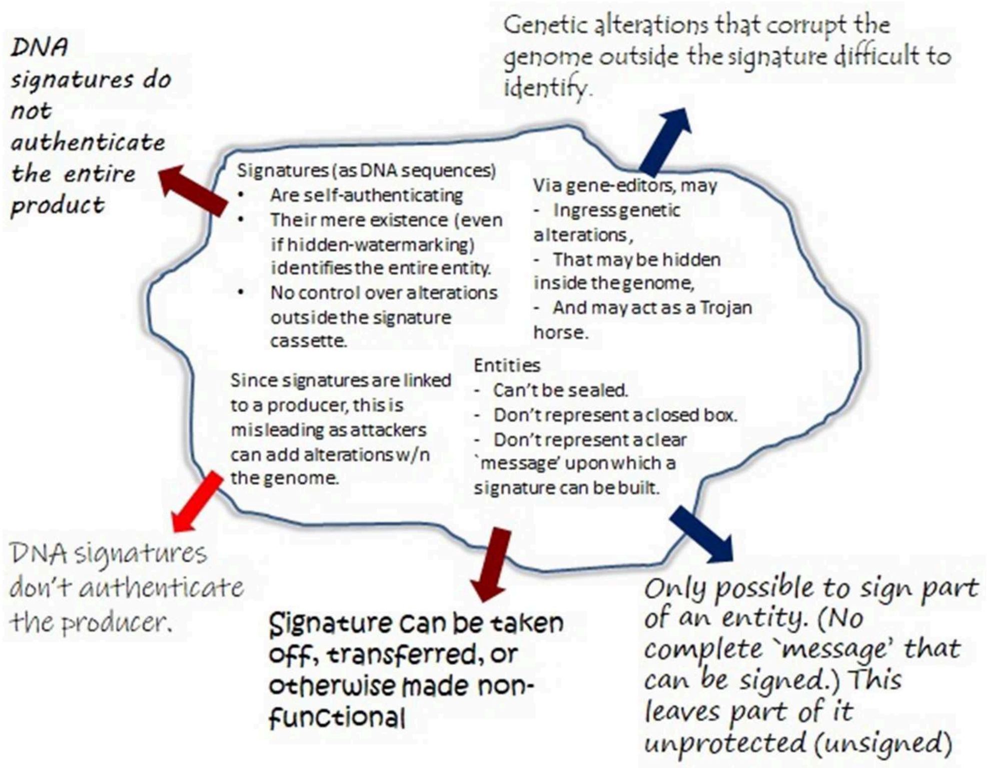

DNA signatures do not authenticate the entire product
Genetic alterations that corrupt the genome outside the signature difficult to identify.
Signatures (as DNA sequences)
• Are self-authenticating
• Their mere existence (even if hidden-watermarking) identifies the entire entity.
• No control over alterations outside the signature cassette.
Via gene-editors, may
- Ingress genetic alterations,
- That may be hidden inside the genome,
- And may act as a Trojan horse.
Entities
- Can't be sealed.
- Don't represent a closed box.
- Don't represent a clear `message' upon which a signature can be built.
Since signatures are linked to a producer, this is misleading as attackers can add alterations w/n the genome.
DNA signatures don't authenticate the producer.
Signature can be taken off, transferred, or otherwise made non-functional
Only possible to sign part of an entity. (No complete `message' that can be signed.) This leaves part of it unprotected (unsigned)

FIGURE 1 | Major shortcomings of DNA signatures compared to traditional and cryptographic signatures. Traditionally, a number of security properties were obtained by sending a message concealed from outside manipulations, in form of sealed envelopes with signatures. This approach helped to ensure integrity (content of the message), its authenticity (sender and receiver), and confidentiality (the content is kept from access and alterations through unauthorized third parties). Similar features can be obtained by cryptographic signatures, by applying a mathematical algorithm ("signing") to some fixed piece of information ("the message"). Importantly, any alterations to "the message" would not only be detected, but would invalidate the signature. The task of signing biologic entities is significantly more complex. This figure summarizes the critical vulnerabilities identified in the text (see section 3.2).

developed by Paracchini et al. (2017) would only identify the specific unauthorized strains disclosed in Europe, but would be of no help in the detection of any modifications that have been introduced in clandestine.

3.2.2. DNA Signatures for the Identification of GMOs

The core vulnerability with DNA signatures is that the mere presence of such signatures is no safeguard against alterations outside the signature cassette. In fact, the very presence of identifiers may establish an effective way to support the hostile usage of GMOs. In Mueller (2019) it is suggested that DNA signatures may intentionally be exploited to enable the counterfeiting of GM-plants that clandestinely have been genetically manipulated to harbor a hazard or other illicit trait. This may involve rather harmless modifications resulting in GMOs authorized in one jurisdiction but not in another. It could also involve much more serious forms of adulterations that turn plants into potential attack vectors, such as the intended deletion or silencing of genes or mechanisms to corrupt the various defense mechanisms employed by plants or to interfere with transgene expression levels in specific plant tissue (see also **Figure 3** below).

Such hazardous GMOs[1] would still bear the unique DNA signatures and thereby give a false sense of security when only these identifiers are tested. Such tests would not reveal the covertly introduced modifications. Unless rigorously analyzed, either through WGS or via phenotype expression patterns, the clandestine manipulation would remain hidden. As the identification of GMOs (especially when dealing with unauthorized ones), is known as practically both difficult and costly (see e.g., Holst-Jensen et al., 2012; Arulandhu et al., 2016; Grohmann et al., 2019), this may further help evil-doers to circulate manipulated products masqueraded as the real thing.

3.3. The Challenge of Authenticating Bioengineered Entities

3.3.1. Issues With Watermarking in Biological Mediums

The problem with DNA signatures is not only illicit adulterations that leave the signature itself unaffected. Of greater concern is the fact *that* with agricultural GMOs those signatures are stably integrated into the genome. Counterfeiters could utilize

[1] Throughout, a hazardous GMO refers to one that intentionally has been manipulated.

this very fact even when the signature string itself is hidden (as it is in watermarking). Modifications that potentially result in a hazardous GMO could lead to serious concerns relating to ownership and attribution especially when the manipulated GMO contains an authentic signature.

Traditionally, watermarking has been an effective means to validate ownership of a certain product. The approach has many useful applications and was utilized, e.g., by Gibson et al. (2010) for the identification of the first artificial bacterium. The incorporation of specific markers into the genome served to identify it as artificially constructed by these researchers. It is important to note that the goal was not to hide a secret. The "hiding" of the identifiers—as with other DNA watermarking methods—was for biocompatibility purposes (i.e., to not interfere with cellular processes); it was not to conceal the existence of these strings from potential signature thieves.

The problem with watermarks for identification purposes is that attackers could potentially misuse their mere presence in form of counterfeiting attacks. All they need to know is *that* a certain GMO is carrying a unique signature (or tag). As long as that sequence is used for identification, this gives a basis for attackers to ingress modification and distribute the adulterated as a counterfeit of the authentic product.

3.4. Signature Transferability Enables Counterfeiting

The above vulnerabilities can be summarized as, (1) the genome carrying the identifier could be corrupted at some other loci, (2) the identifiers themselves could be transferred unto an unauthorized entity, or (3) they could be duplicated (stolen), to masquerade an adulterated product as the original. Thereby, the manipulated GMO:

- Is assumed to come from a claimed authentic producer,
- And thus is believed to resemble the claimed product.

It is seen that this constitutes a novel form of counterfeiting in essentially two ways. The first is to engineer a product that is cheaper to produce, cultivate, or select for (such as through the unauthorized use of antibiotic markers). Of greater concern is counterfeiting with a malicious intent, when trying to falsely attribute an adulterated product to a legitimate manufacturer. These two situations need to be addressed differently. In the former case it is paramount to keep counterfeiters from introducing manipulated GMOs (which could contain hazardous modifications) into the market. The latter concerns adulterated GMOs that attackers have already managed to bring into circulation, and that require adequate methods to assign attribution. A summary of this, along with critical requirements for a solution, is depicted in **Figure 2**.

4. PRACTICAL CONSIDERATIONS AND PRELIMINARY RECOMMENDATIONS

4.1. Cryptographic Primitives in the Context of Bioengineering

Cyber-security risks are often illustrated in form of a network depicting the insecure channels between users Alice and Bob. For instance, if Alice is sending an email to Bob, then this is a complex process, whereby the email message is broken down into parts, processed at various points in this communication system, until it eventually reaches Bob. The individual steps are all computers or processes, and notably, each of these can become the source of attack.

However, in the cyberbiosecurity realm, this picture is radically different. If we consider the production of a GMO, then the individual steps between this (honest) intent (or, alternatively, the goal for the willful alteration of a GMO) to the final product on the market also consists of numerous steps. Yet, in sharp contrast to the cyber-realm, here it is not the case that each and any individual node in the network is more or less equally equipped for introducing the same degree of harm.

Although the - minimal - requirements that are needed by bio-cyber hackers to misuse modern gene-editors such as CRISPR/Cas have been highlighted by many (see e.g., DiEuliis and Giordano, 2017; Dunlap and Pauwels, 2017), this does not address the feasibility and likelihood of performing attacks on entire GMOs - to the extent that these would have a noticeable phenotypic effect at a large scale.

To achieve a large-scale hazardous effect, more is required than just manipulating a few cells of a target organisms (which may even be within a secured physical environment). This leads to considerable operational, research, and manufacturing challenges. It is seen that the most critical factor lies in the actual abilities of attackers, to generate and distribute their fabricated GMOs, while evading existing screening and safety checks. It is suggested here that this establishes a scale of attack feasibility that needs to be overlaid with the previous (cryptography based) factors, to estimate the practical likelihood of intrusions.

Figure 3 describes such a hierarchy in the context of trying to weaponize agricultural GMOs (see Mueller, 2019).

4.2. Key Approach

What **Figure 3** (with special focus on GM plants Mueller, 2019) shows, is the difference of the influence of the various cryptographic primitives on various levels of attack. At the lower end of the hierarchy we find attackers that can do simple gene-edits, but that don't rely on the manufacturing facility itself to produce these changes. In contrast, the top describes attackers that are able to essentially sabotage an entire GMO, or even an entire GMO production facility.

As section 2.3.1 illustrates, many cyberbiosecurity channels based on the defining characteristics of integrity are in fact insecureable. On the other hand, section 3.2 describes critical vulnerabilities in form of counterfeiting attacks. The latter are based on compromising the product itself (corrupting integrity) as well as authenticity (who generated the product). Although restoring integrity seems to be difficult (if not impossible, see above), it may be possible to strengthen channels based on authenticity and confidentiality. Indeed, as **Figure 3** shows, it is the scale of compromising these (involving both digital and physical/biological entities) that determines the severity of attacks (see also **Figure 1**, where red arrows are associated with authenticity/confidentiality, and blue with integrity).

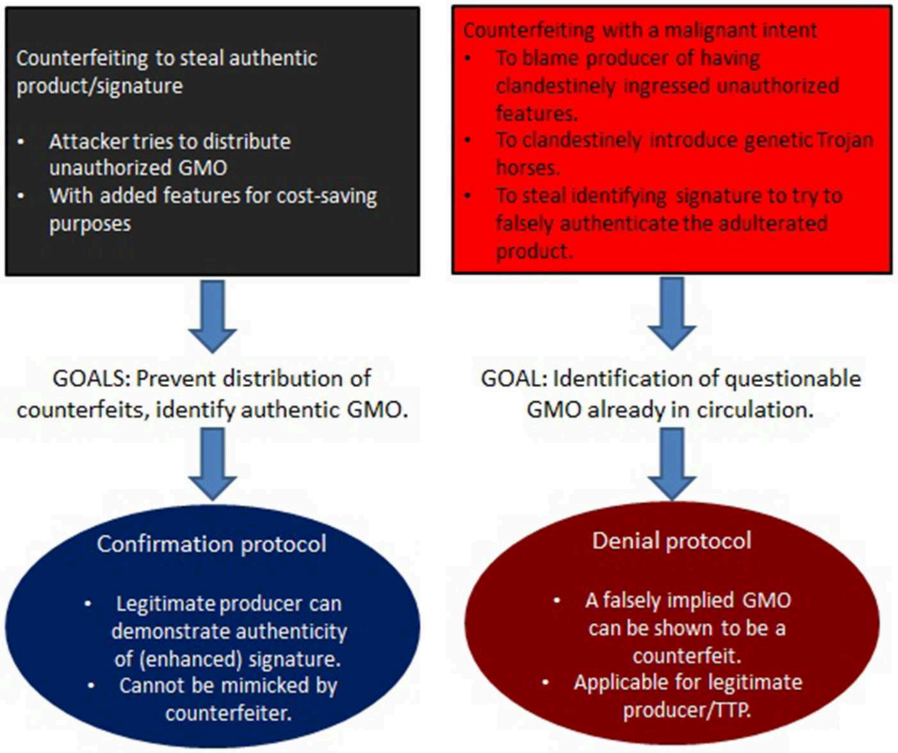

Counterfeiting to steal authentic product/signature
• Attacker tries to distribute unauthorized GMO
• With added features for cost-saving purposes
Counterfeiting with a malignant intent
• To blame producer of having clandestinely ingressed unauthorized features.
• To clandestinely introduce genetic Trojan horses.
• To steal identifying signature to try to falsely authenticate the adulterated product.
GOALS: Prevent distribution of counterfeits, identify authentic GMO.
GOAL: Identification of questionable GMO already in circulation.
Confirmation protocol
• Legitimate producer can demonstrate authenticity of (enhanced) signature.
• Cannot be mimicked by counterfeiter.
Denial protocol
• A falsely implied GMO can be shown to be a counterfeit.
• Applicable for legitimate producer/TTP.

FIGURE 2 | Herein, unrecognized risks involving counterfeiting attacks are identified that rely on the intentional misuse of prevailing DNA signatures (section 3). Although no such GMO counterfeit is confirmed in circulation, a recent B2 contamination event in Europe (Paracchini et al., 2017) demonstrates that these risks need to be taken seriously. Depending on the type of risk, different strategies need to be pursued. Steps toward realizing these goals are described in sections 4, 5.

This suggests the following approach to help mitigate the risks associated with DNA signatures as described above. Essentially, the goal is to keep the product (especially, the signature) from all those intended to have it. This begins with the producer and proceeds along the supply chain. While several approaches have been suggested by Frazar et al. (2017) to help secure the synthetic biology supply chain, here an enhanced concept of DNA signatures is suggested to help support this, as follows. Such signatures should:

- Incorporate a strengthened feature to enable authentication of the producer. As continuous quality control testing of the production process and product purity is required to exclude contaminations and/or impurities (Hermann and Schurter, 1995; Paracchini et al., 2017), ensuring the origin of the GMO (authenticity) is most critical to secure integrity of the released product.
- Ensure confidentiality–of the enhanced signature (to prevent signature theft and transferability). That is, keep the identifier (signature) from all but those authorized to have it.
- Allow public verifiability of GMOs for the verification of authentic products.

These last two items seem to be conflicting. Yet, it will be shown below that they can be reconciled by realizing DNA signatures in two parts, one digital, and one physical.

5. A CRYPTOGRAPHY-BASED METHOD TO ENHANCE DNA SIGNATURES

5.1. Intuitive Description of Key Features

The presented solution relies on a cryptographic mechanism which had been constructed to address a challenge that arose in a completely different context. An undesirable characteristic of digital signatures is that anyone who has access to the deciphering key (in the "public key" setting this would be everyone) would be able to verify the validity of a purported signature string. This universal verifiability (or self-authentication) would be unacceptable when sensitive or private information is involved. A typical example is described by Xia (2013), when software vendors might want to sign on their products to provide authenticity to their paying customers. At the same time, however, they do not want those who have illegally duplicated their software to verify the validity of the product by being able to verify their signature.

This situation is similar to the one described above (section 3.2). DNA-signatures have been constructed so that they can be publicly verified. While this is a critical factor to support the identification of GMOs, this may also enable their potential misuse. An attacker can introduce changes into the genome without affecting the authenticating signature identifiers, allowing the adulterated product to pass as the real thing. This

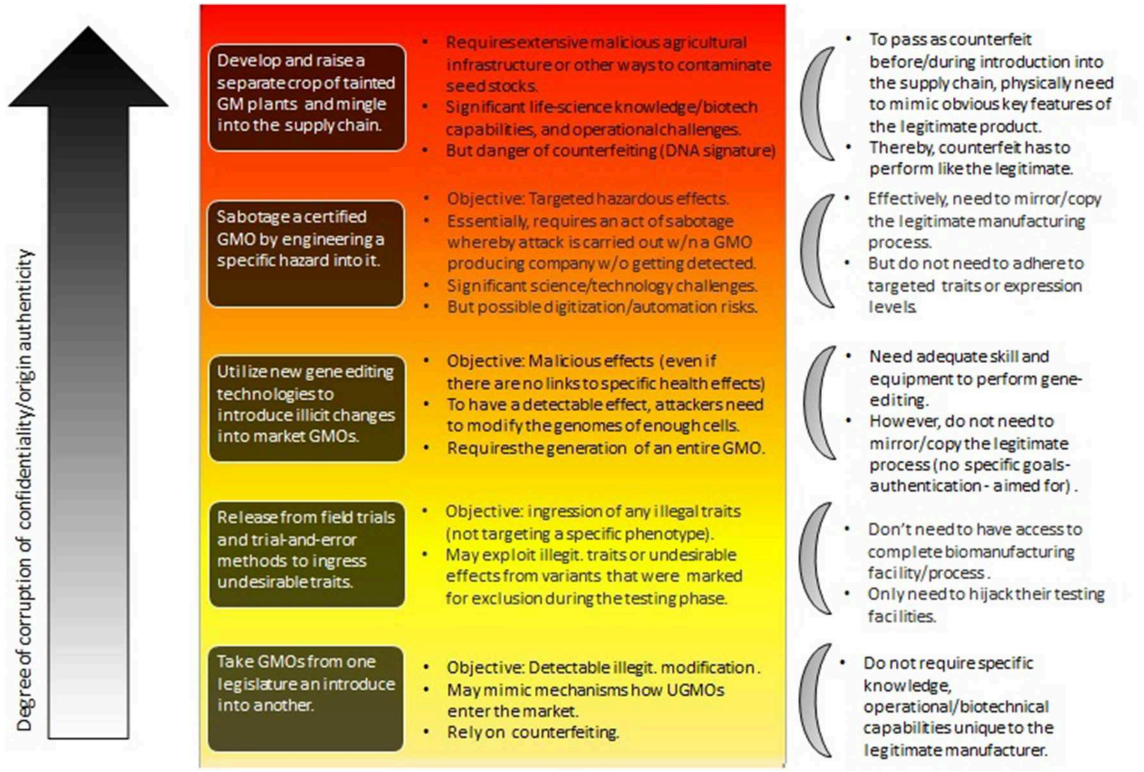

Degree of corruption of confidentiality/origin authenticity
Develop and raise a separate crop of tainted GM plants and mingle into the supply chain.
• Requires extensive malicious agricultural infrastructure or other ways to contaminate seed stocks.
• Significant life-science knowledge/biotech capabilities, and operational challenges.
• But danger of counterfeiting (DNA signature)
• To pass as counterfeit before/during introduction into the supply chain, physically need to mimic obvious key features of the legitimate product.
• Thereby, counterfeit has to perform like the legitimate.
Sabotage a certified GMO by engineering a specific hazard into it.
• Objective: Targeted hazardous effects.
• Essentially, requires an act of sabotage whereby attack is carried out w/n a GMO producing company w/o getting detected.
• Significant science/technology challenges.
• But possible digitization/automation risks.
• Effectively, need to mirror/copy the legitimate manufacturing process.
• But do not need to adhere to targeted traits or expression levels.
Utilize new gene editing technologies to introduce illicit changes into market GMOs.
• Objective: Malicious effects (even if there are no links to specific health effects)
• To have a detectable effect, attackers need to modify the genomes of enough cells.
• Requires the generation of an entire GMO.
• Need adequate skill and equipment to perform gene-editing.
• However, do not need to mirror/copy the legitimate process (no specific goals-authentication - aimed for).
Release from field trials and trial-and-error methods to ingress undesirable traits.
• Objective: ingression of any illegal traits (not targeting a specific phenotype).
• May exploit illegit. traits or undesirable effects from variants that were marked for exclusion during the testing phase.
• Don't need to have access to complete biomanufacturing facility/process.
• Only need to hijack their testing facilities.
Take GMOs from one legislature an introduce into another.
• Objective: Detectable illegit. modification.
• May mimic mechanisms how UGMOs enter the market.
• Rely on counterfeiting.
• Do not require specific knowledge, operational/biotechnical capabilities unique to the legitimate manufacturer.

FIGURE 3 | The types of attacks involving GM plants as considered by Mueller (2019) (central part of the figure), roughly ordered from bottom to top relative to their risk-potential. Their impact is also hierarchical with risks at the lower level inherited at higher-levels. Herein, the focus is on the degree to which confidentiality and authenticity are violated (see section 4.2).

is of particular concern if only the signatures are used for identification. They would indeed be universally verifiable, but would not say anything about any hidden adulterations within the genome.

The predicament of self-authenticity of digital signatures motivated the introduction of undeniable signatures (Chaum and Van Antwerpen, 1990), designated confirmer signatures (Chaum, 1995; Camenisch and Michels, 2000), and improvements (see e.g., Camenisch and Michels, 2000; Ateniese, 2004; El Aimani, 2009; Xia, 2013). Essentially, their realization hinges upon two main features as summarized in **Figure 4**. The first involves Zero-Knowledge (ZN) proofs of knowledge which allow (mathematical) demonstration of knowledge of the signer's secret identifier without ever having to expose this secret. The second is called "invisibility" (Galbraith and Mao, 2003) and in this context means that outsiders would not be able to distinguish between two types of secrets. This concept is critical for ensuring a form of authenticity (unimpersonation, see **Figure 4** and Xia, 2013) and will be incorporated below in two ways. When invisibility is combined with ZK (**Figure 4**), then this also supports confidentiality - the secret is kept from all but those authorized to have it. This cryptographic framework, when applied to enhance DNA signatures, would address both of the key goals identified above (section 4.2). The involvement of the most critical elements is detailed in **Figure 4**.

5.2. Summary of the Enhanced DNA Signature Method

Designated confirmer signatures and ZK proofs were first suggested in Mueller (2014), Mueller et al. (2016) as a basis to mitigate the problem of DNA signature transferability described above. These authors also provided an explicit description of algorithms and a specific watermarking protocol to hide a representation of the digital signature component within the GMO itself.

As the underlying cryptographic framework has been summarized and enhanced by Xia (2013), this now allows for a more direct description of such DNA signatures, that allows additional improvements relative to Mueller et al. (2016). Thereby, enhanced DNA signatures can be based on two legs (via a cryptographic/digital protocol, and via DNA bases/physically), with the following main features.

- Signature strings are not self-authenticating in the sense that they can directly be verified (simply by their physical or electronic existence).
- Instead, signature verification is firstly done via a cryptographic process as summarized in section 5.1 and explained more fully in section 5.3. Thus, it is the outcome of that process that determines the conclusion, not the digital (or physical) signature sequence itself.

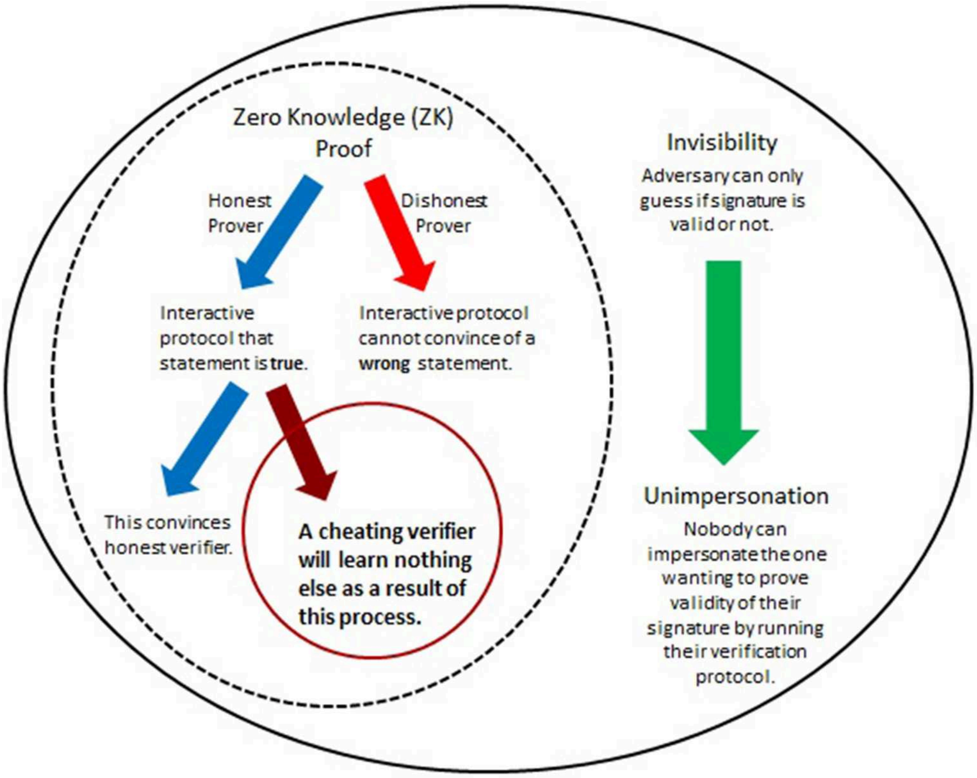

Zero Knowledge (ZK) Proof
Honest Prover
Dishonest Prover
Interactive protocol that statement is true.
Interactive protocol cannot convince of a wrong statement.
This convinces honest verifier.
A cheating verifier will learn nothing else as a result of this process.
Invisibility
Adversary can only guess if signature is valid or not.
Unimpersonation
Nobody can impersonate the one wanting to prove validity of their signature by running their verification protocol.

FIGURE 4 | Herein, improvements of DNA signatures are obtained by utilizing cryptographic tricks that have proven useful for special cryptographic applications such as identification protocols and enhanced signatures (Menezes et al., 1996; Camenisch and Michels, 2000; Ateniese, 2004; El Aimani, 2009; Xia, 2013). At the core are (mathematical) interactive proof systems to demonstrate the (in)validity of a certain statement such as, "This is my personal PIN." The significance of Zero-Knowledge (ZK) proofs lies in the fact that such systems can convince of the correctness of the statement without needing the involved parties to expose any details, such as, specifics of the PIN itself. ZK protocols can be overlaid with a feature that ensures authenticity of the originator of the statement or signature. When combined, this gives a powerful method to verify signatures while at the same time preventing their transferability or misuse by unauthorized parties.

- Secondly, a "cryptographic fingerprint" (hash) of the cryptographic signature is converted into DNA bases via some watermarking protocol and incorporated into the genome (see section 5.4 for details).
  - In case of confirmation of a GMO (**Figure 2**), the presence of these DNA sequences can be publicly verified by standard hybridization methods (This part in itself incorporates no security components and serves only to tie the digital to the physical component).
  - To achieve denial of a signature (**Figure 2**), the above is complemented by a physical invisibility feature (**Figure 4** and section 5.1). In case of dispute, when denial is required, this step can be performed by competent authorities in combination with WGS.

For the digital part, the objective of the signer (the producer of a GMO) is to convince a buyer (e.g., key importers) of the validity of the cryptographic signature (as described in sections 5.1, 5.3). Yet, the signature string itself is not exposed during this process. Its involvement is implicitly, that is, in a hidden manner (ZK property, **Figure 4**). This allows it to remain concealed throughout. The physical part is primarily used for linking the digital with the actual product, but also serves to solve dispute. A summary of the method is given in **Figure 5**.

5.3. The Digital/Cryptographic Part

This section summarizes the security properties of the digital part in the framework developed above. For an explicit formulation via specific cryptographic algorithms, parameters and keys involved, see (Mueller, 2014; Mueller et al., 2016). The most important components to enhance DNA signatures are the following:

- The cryptographic verification process can be run by legitimate parties (most notably, the manufacturer and the first point of sale, Golan et al. (2004) in form of a series of check routines. The completion of these interactions establishes a mathematical proof which convinces the verifier about the (in)validity of the purported signature, including its alleged originator (authenticity).
- Nobody can see the validity of a signature without the verification protocol. In other words, an adversary who has access to a purported signature string has no choice other

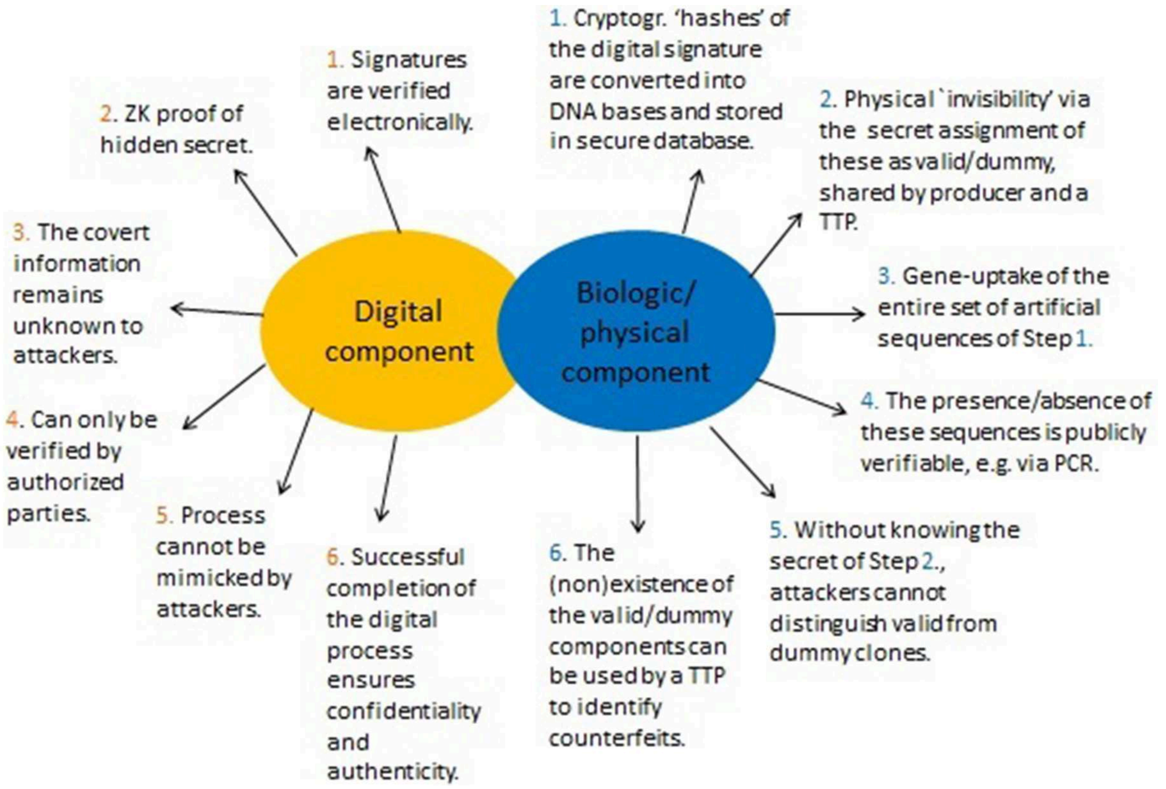

1. Signatures are verified electronically.
2. ZK proof of hidden secret.
3. The covert information remains unknown to attackers.
4. Can only be verified by authorized parties.
5. Process cannot be mimicked by attackers.
6. Successful completion of the digital process ensures confidentiality and authenticity.
Digital component
Biologic/ physical component
1. Cryptogr. 'hashes' of the digital signature are converted into DNA bases and stored in secure database.
2. Physical `invisibility' via the secret assignment of these as valid/dummy, shared by producer and a TTP.
3. Gene-uptake of the entire set of artificial sequences of Step 1.
4. The presence/absence of these sequences is publicly verifiable, e.g. via PCR.
5. Without knowing the secret of Step 2., attackers cannot distinguish valid from dummy clones.
6. The (non)existence of the valid/dummy components can be used by a TTP to identify counterfeits.

FIGURE 5 | Summary of the proposed method to enhance DNA signatures. Signatures are represented and verified in two ways. One is digital and based on specific cryptographic signatures (section 5.1) by utilizing enhanced Zero-Knowledge (ZK) proofs of knowledge via a cryptographic "invisibility" property (**Figure 4**). The second part ties the actual (physical) GMO to the digital part and adds a physical "invisibility" feature. Consequently, it is possible to (1) Demonstrate genuineness of a legitimate signature (this can be done both physically and digitally), (2) Prevent counterfeiters from selling manipulated GMOs, and (3) Allows authentic producers to demonstrate that a falsely attributed (fabricated) GMO is not theirs. This step may require WGS and can only be performed by a TTP or competent enforcement authorities who can verify the secret assignment into "valid" or "dummy".

than a random guess to learn if that signature is valid or not (invisibility).

- The conclusion whether the digital string represents a valid signature or not cannot be misused by adversaries (who might be masquerading as verifiers) to impersonate the producer (signature theft). The argument for this is simple. If an adversary can impersonate the legitimate producer (of the signature) by successfully convincing any verifier thereof, then it must be able to run the interactive protocol. However, if an attacker can run the protocol in such a manner to convince a verifier, it must trivially know the signature's validity. In other words, if an attacker can break unimpersonation, it can also break invisibility (**Figure 4**).
- Even if the interactive protocol has successfully convinced the interacting parties of the (in)validity of a signature, attackers (masquerading as verifiers) cannot use any of the insights learned from the underlying mathematical procedures to demonstrate this same fact to anyone else (ZK property).

Consequently, attackers trying to masquerade a legitimate GMO cannot conduct the necessary digital confirmation protocol (**Figure 2**), which will make it impossible to sell their counterfeits. The important point is that only those in possession of a the secret (as required in the cryptographic protocol) are able to complete this process. Thereby, only the legitimate manufacturer of the authorized GMO can provide a digital proof of their genuineness. This step is meant to prevent introduction of unauthorized (and possibly hazardous) GMOs into the supply chain at a larger scale (see **Figure 3**).

The digital part gives assurance that the GMO in question was indeed produced by the legitimate manufacturers, and that it is not a manipulated product of unauthorized origin. While digital verification in many regards may be easier than physical, it needs to be stressed that this part only gives assurance about the purported GMO. Clearly, there needs to be a link to tie the above to the real, physical entity in question. This accomplished by the second leg of the protocol (section 5.4 below).

Overall, the cryptographic component ensures confidentiality and authenticity and thereby helps to prevent the unauthorized distribution of counterfeits as well as the identification of false signatures in case of concern or dispute. Assuming that the production company has verified the authenticity of the final product before its distribution, this guarantees a high degree of security of DNA signatures. See **Figure 6** for a summary of this part.

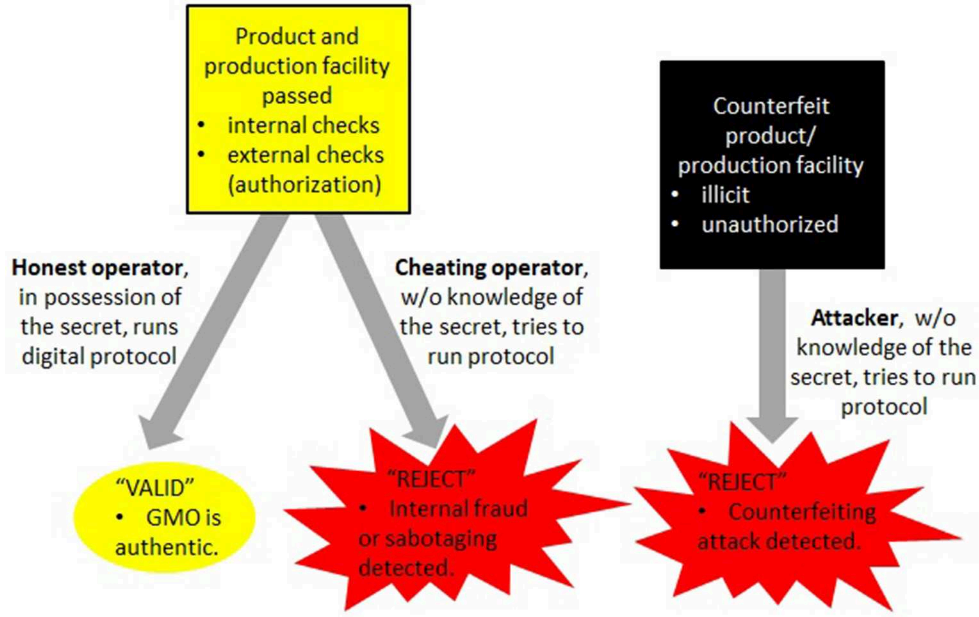

Product and production facility passed
• internal checks
• external checks (authorization)
Counterfeit product/ production facility
• illicit
• unauthorized
Honest operator, in possession of the secret, runs digital protocol
Cheating operator, w/o knowledge of the secret, tries to run protocol
Attacker, w/o knowledge of the secret, tries to run protocol
"VALID"
• GMO is authentic.
"REJECT"
• Internal fraud or sabotaging detected.
"REJECT"
• Counterfeiting attack detected.

FIGURE 6 | The digital part of the enhanced DNA signature method utilizes special cryptographic signatures (sections 5.1, 5.3) whereby signature verification is accomplished via a protocol rather than verification of presence or absence of a certain sequence. This gives a high degree of security and can only be achieved by legitimate producers (or their proxies) who know the underlying secret used for computing these cryptographic signatures. Attackers are not able to mimic this process and therefore cannot distribute counterfeits of GMOs by trying to masquerade them as the original product (**Figure 2**).

5.4. The Physical Component—Tying the Cryptographic Part to the Actual GMO

To link the digital part with the actual GMO, it is necessary to incorporate a representation of it within the genome itself. The strong security properties of the above digital part are depicted in **Figure 4**. Ideally, one might try to extend these attributes to the physical domain (the actual GMO in question). A quick reflection immediately reveals major challenges. Whenever a purported DNA sequence is verified physically, e.g., via hybridization methods, then this obviously reveals the presence or absence of this sequence in the GMO, making it impossible to achieve a ZK property (the solid circle in **Figure 4**). This section offers some suggestions to reclaim related security features nonetheless.

5.4.1. Construction of the Physical Signature Component

A standard characteristic of cryptographic signatures is that their representation in the electronic signature space looks like a random string with equal occurrences of $0's$ and $1's$. However, the nucleotides within the genome do not represent an equidistribution of A,T, C and G. In Mueller et al. (2016) a method is described how the cryptographic signature can be converted into the DNA alphabet so that it is indistinguishable from endogenous DNA after insertion into the genome.

The watermarking protocol that was developed in Mueller (2014) and Mueller et al. (2016) for this purpose takes advantage of the equiprobable distribution of the cryptographic (usually, binary) alphabet, to represent binary text triplets according to the codon bias of the host genome (Figure 3 in Mueller et al., 2016). Doing this effectively camouflages the resulting DNA string so that an adversary cannot easily identify its presence within the genome (other than through WGS). In Mueller et al. (2016) this physical signature is only required in case of conflict and normally the signature remains hidden.

This approach can be enhanced, in two ways. The first is to shorten the DNA signature, so that in place of the rather lengthy cryptographic signature, only a cryptographic fingerprint (hash) (see e.g., Menezes et al., 1996) is incorporated into the genome. Such hashes have the beneficial property that they are much shorter (a few hundred bases). Yet, in practice this is sufficient, as it is infeasible to retrieve any useful information about the signature that the hash was computed from. The second improvement will make the absence or presence of DNA signature sequences publicly verifiable (section 4.2), and thereby enable identification of authentic GMOs in general circumstances, and not only during dispute.

To accomplish this, a form of "invisibility" property (**Figure 4**) will be achieved within the DNA signature part itself, as follows.

First, several different hashes are created from the given cryptographic signature and stored in a secured and publicly accessible database. In a second step, these are secretly assigned different values, denoted "valid" and "dummy." This assignment of which string is of which type is secretly shared between the producer and a Trusted Third Party (TTP). The entire set of the hashes are then converted into DNA bases as in Mueller et al. (2016) and embedded into various clones.

It is important to note that since the signature components within the genome (that is, the individual hashes) are meant

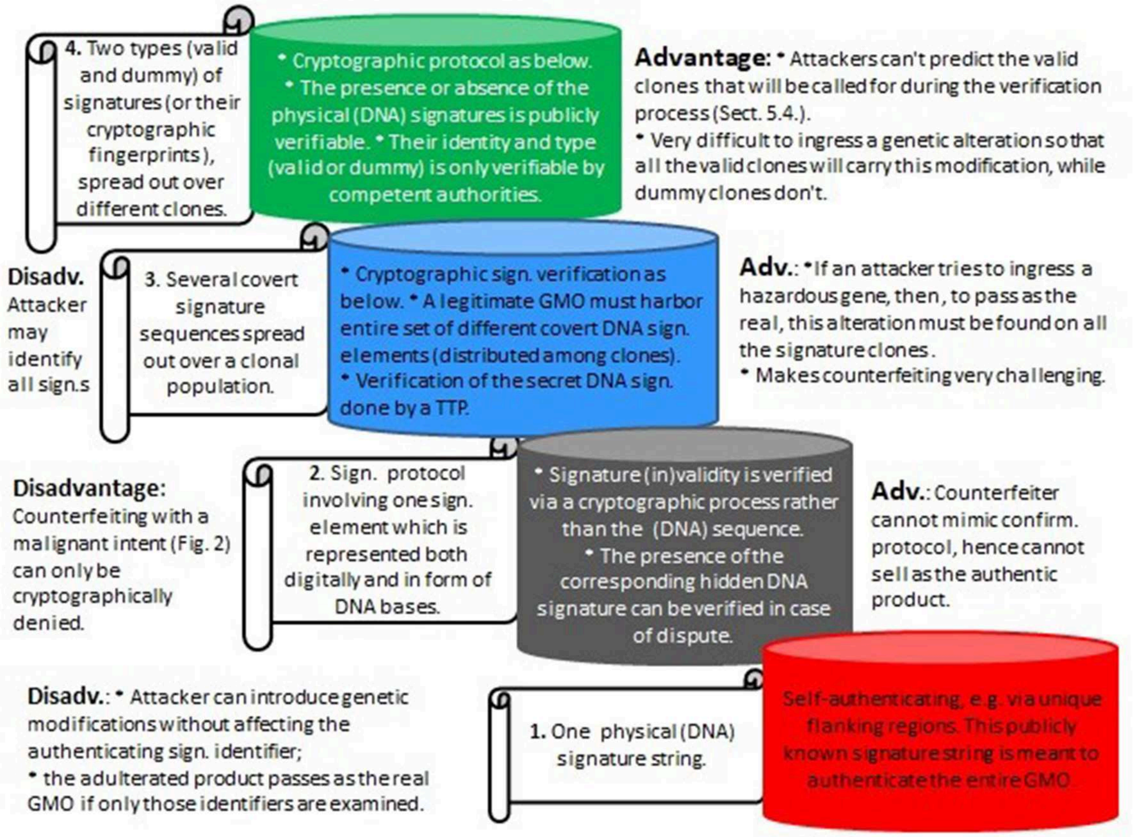

4. Two types (valid and dummy) of signatures (or their cryptographic fingerprints), spread out over different clones.
• Cryptographic protocol as below. • The presence or absence of the physical (DNA) signatures is publicly verifiable. • Their identity and type (valid or dummy) is only verifiable by competent authorities.
Advantage: • Attackers can't predict the valid clones that will be called for during the verification process (Sect. 5.4.).
• Very difficult to ingress a genetic alteration so that all the valid clones will carry this modification, while dummy clones don't.
Disadv. Attacker may identify all sign.s
3. Several covert signature sequences spread out over a clonal population.
• Cryptographic sign. verification as below. • A legitimate GMO must harbor entire set of different covert DNA sign. elements (distributed among clones).
• Verification of the secret DNA sign. done by a TTP.
Adv.: • If an attacker tries to ingress a hazardous gene, then, to pass as the real, this alteration must be found on all the signature clones.
• Makes counterfeiting very challenging.
Disadvantage: Counterfeiting with a malignant intent (Fig. 2) can only be cryptographically denied.
2. Sign. protocol involving one sign. element which is represented both digitally and in form of DNA bases.
• Signature (in)validity is verified via a cryptographic process rather than the (DNA) sequence.
• The presence of the corresponding hidden DNA signature can be verified in case of dispute.
Adv.: Counterfeiter cannot mimic confirm. protocol, hence cannot sell as the authentic product.
Disadv.: • Attacker can introduce genetic modifications without affecting the authenticating sign. identifier;
• the adulterated product passes as the real GMO if only those identifiers are examined.
1. One physical (DNA) signature string.
Self-authenticating, e.g. via unique flanking regions. This publicly known signature string is meant to authenticate the entire GMO.

FIGURE 7 | Various types of DNA signatures as considered herein, from bottom to top with increasing levels of security. 1. Represents the existing DNA signature paradigm (e.g., Levine, 2004); 2. and 3. are described in Mueller et al. (2016), and 4. (section 5) is an extension of the cryptographic invisibility feature which is central to the underlying cryptographic part in Mueller et al. (2016). (Sign, Signature; Adv, Advantage; Disadv, Disadvantage; Confirm, Confirmation).

to be publicly verifiable, that here the watermarking protocol is not employed to hide the sequences from adversaries, but mainly for biocompatibility and practical purposes. This step may be complemented by various techniques to support the uptake of artificial sequences, relative to GC content, codon bias, repetitive sequences etc. Practical steps how the identification of the individual target sequences (e.g., through hybridization) via suitable primers can be aligned with related methods have been described elsewhere (e.g., Paracchini et al., 2017).

5.4.2. Detection of an Authentic GMO (Confirmation)

Because the secret lies in the designation of the various clones as valid or dummy, the individual DNA signature hashes (without this assignment) can now serve just as previous DNA signature sequences. As artificially created constructs (hashes of the cryptographic signature), it is unlikely that these overlap with endogenous sequences of the GMO (although this could be verified a priori). As a result, their presence or absence not only authenticates a unique GMO, but also establishes a verifiable link to the cryptographic protocol (which holds the core of the security qualities of the entire protocol).

5.4.3. Confirmation of a Counterfeit (Denial of an Unauthorized GMO)

The secret association of which clones are carrying valid and which mere dummy elements can also help resolve the following vulnerability not addressed in Mueller et al. (2016). Suppose an adversary (masquerading as an honest buyer) interacts with the manufacturer or their proxy in the cryptographic protocol, but manages to adulterate the authenticated GMO (see **Figure 3** for practical difficulties). Rather than trying to steal a product, an attacker could try to blame a producer for creating such (possibly harmful) modified GMOs, and could try to support their claim through the cryptographic protocol associated with the verification of an authentic GMO (**Figure 2**).

Although the cryptographic protocol can be used by honest parties to digitally "deny" a falsely attributed signature as theirs, in the present context this is absurd. What is at stake here is not whether or not GMO manufacturers can mathematically prove that a cryptographic signature is not theirs, but that the actual (manipulated) GMO was *not* manufactured by them. This can be achieved through the secret assignment of the individual clones as valid or dummy (physical invisibility). Thus, as in the digital part, without knowledge of their secret underlying meaning (as valid or dummy), outsiders can only speculate which clone is of which type. The above vulnerability can now be resolved by a TTP or competent authority according to the following rules.

- Legitimate GMOs can be identified if the collection of clones contain the entire set of signature hashes. A collection of clones that does not include the full set of signatures is unauthentic and possibly the result of an attack.
- By definition, signature verification only involves clones carrying valid signature components. That is, as legitimate participants can distinguish the two types of clones (associated to "valid" or "dummy") it is possible to select only those samples with the former type of signatures.
- Also by agreement, any genetic modifications (used by the attacker to blame the manufacturer, or identified via WGS) found on "dummy" clones are declared counterfeits.

The reasoning for this is that legitimate manufacturers will not introduce genetic modifications other than those they are seeking (or owning) authorization for. Thus, they will not include unauthorized modifications to any of their clones. Any additional alterations, e.g., for testing purposes, can be forced upon valid clones only, which however requires knowledge of this secret (see also section 5.4.4 below).

One of the critical goals of microbial forensics lies in the identification of the causative agent or source of a disease outbreak (Murch, 2015). This is equally important with illicit or compromised GMOs. The manipulated and masqueraded product could be mingled into legitimate supply chain and widely distributed. In theory, it seems impossible to avoid such intrusions, at least on a local level. Yet, as with microbial forensics, "resolving with high confidence whether or not the outbreak" - or the occurrence of illicit or manipulated GMOs - "manifested as a result of a natural, accidental, or deliberate event is crucial" (Murch, 2015). While more difficult to realize, the above denial part may effectively complement previous efforts to assign molecular attribution (see also Minogue et al., 2019).

5.4.4. Summary and Extensions of the Denial Part

When the enhanced cryptographic signatures based on ZK proofs were first invented (Chaum and Van Antwerpen, 1990; Chaum, 1995), they were realized via very simple mathematical algorithms such as simple modular arithmetic (Menezes et al., 1996). It quickly became apparent that verifying a secret via a proof of knowledge is considerably more difficult to realize in the case of denial. Obviously, it would not be enough to just claim to not know the secret. Cryptographic realizations to support a denial feature have required somewhat more sophisticated mathematical algorithms, which may explain why these types of signatures have not received much attention.

Importantly, this same challenge of denial can effectively be extended into the physical realm. This is the essence of the denial protocol described above. Here, the role of the secret is assumed by the value of the clones (valid or dummy). Now, in form of a physical test, the challenge of demonstrating knowledge (or its opposite in form of denial) can easily be realized through hybridization methods that verify the presence or absence of the required sequences.

Assuming that the artificial signatures are stably integrated into the genome, the absence of some "valid" clones will immediately point to an attack. However, much more problematic is the situation where skilled attackers could aim to circumvent that. With modern gene editors it might be possible to illegally ingress genetic hazards into the complete set of all the (different) authentic clones. As a result, the corrupted set of GMOs would carry the required set of signature elements, which could lead to critical problems concerning legal ownership and attribution. It is for reasons like this that the proposed method requires not only different signature clones, but also the added invisibility component. See **Figure 7** for the different signature types considered herein and their advantages and disadvantages.

The denial part is the only part in the proposed method that relies on WGS (unless knowledge and assumed functionality of the adulterations are already part of the attack). This is necessary to identify all genetic modifications that clandestinely have been inserted into the GMO. While this is costly, WGS continues to become more practical and efficient (see e.g., Arulandhu et al., 2016). Moreover, the search could be enhanced by biochemical identification systems as was demonstrated by Paracchini et al. (2017) in the characterization of the unauthorized GM *B. subtilis* strain.

Based on their in-depth analysis, these authors came to the conclusion that some of these alterations were intentionally introduced. Nonetheless, at this point one can only speculate whether such manipulations were done as part of an actual criminal act (**Figure 2**). The denial component above would give authorities a method to demonstrate such types of counterfeiting attacks. Since legally authorized parties can obtain the secret designation of the type of the clones (as valid or dummy), it is possible for them to amplify and select only the valid clones and verify the presence/absence of the genetic alteration on the individual clones. As any illegal modification after approval of the authorized GMO will affect large proportions of clones, inadvertently some of these alterations will land on clones that were labeled as "dummy." This would officially confirm that the alteration is a form of attack.

However, not all attacks need to involve genetic modifications (see **Figure 3**). Additional known risks involve the giving away of GMOs from laboratory or field trials. Of special concern here are trial-and-error methods (e.g., Rodríguez-Leal et al., 2017), whereby inside-attackers can take seeds or products exhibiting undesirable ("the error") characteristics.

Here, the secret classification of the various clones can be helpful in the following way. Laboratories seeking to run a certain experiment can exchange the secret identifier information with a trusted party or certification authority. Thereby, they would agree to perform the trial-and-error methods only on a certain subset of the different clones. Insider attackers that are injecting "error" samples into the food/supply chain would inadvertently select from this same specific set of clones, demonstrating that the illicit sample needed to have originated from their lab, thus narrowing down possible suspects - and perhaps discouraging insiders to perform such types of illegal actions.

TABLE 4 | Summary of the proposed method to enhance DNA signatures, relative to the two main goals of disputing a falsely alleged GMO or confirming a genuine one (see **Figure 2**).

Performance	Meaning and significance	As realized in the proposed protocol
True positive	An authentic GMO can be verified as such. Signature verification protocol returns "ok"	1. Manufacturer/proxy can successfully run the cryptographic confirmation protocol 2. The existence of the hashes of all the signature transgenes within the GMO is publicly verifiable (e.g., via hybridization)
False positive	The protocol falsely identifies/approves an unauthorized/adulterated GMO (danger of distributing a counterfeit)	Not possible, due to the cryptographic part of the protocol (as a necessary requirement to bring GMOs into circulation) : By virtue of the ZK property, attackers cannot impersonate true manufacturer; hence, cannot sell a counterfeit. The digital part is linked to the physical via signature hashes
True negative	An unauthorized GMO can be confirmed as such. Important that this is done via the physical part of the protocol as the digital part only gives information *about* the object	Physical denial part. Thereby, a GMO is *not* authentic, if at least OFTF: 1. Not the complete set of signature hashes present within the genome (publicly verifiable via PCR, etc.) 2. Verification by competent authorities (who have access to the secret of which clones are valid/dummy), according to the following a. Identify genetic adulterations (may require WGS) b. Amplify all valid clones c. If the illicit genetic alteration is found on a dummy clone, the GMO is a counterfeit
False negative	A genuine GMO is identified as inauthentic	Not possible, due to (1) the correctness/completeness of the digital part (an honest prover can successfully run the protocol), and (2) as long as physical signature components within the genome are stably integrated

5.5. Summary of the Overall Protocol

An overall summary of the enhanced DNA signature method is given in **Figure 5**. It is seen that by incorporating the fundamental steps of special ZK based cryptographic signatures (**Figure 4**), it is possible to obtain the necessary safeguarding components as identified in section 4.2.

The strength of the described method lies in its robust identification of GMOs which cannot be mimicked by counterfeiters. For practical purposes, this may indeed be one of the most important steps to guarantee the circulation of authentic GMOs. The opposite (more costly) part of denial is practically of less concern and in fact is already minimized when authenticity and confidentiality of GMOs are ensured. **Table 4** gives a survey of these opposing goals and the components of the solution as realized herein.

Although the method by Mueller et al. (2016) (that part of the cryptographic solution presented herein is based on) focused on GM plants, it can be extended to different GMOs, provided the transgenic cassette can be stably integrated into the host genome. Especially with bacteria, special focus needs to be placed on that, as artificial sequences not integrated into the bacterial genome but onto extra-chromosomal plasmids may be lost (see e.g., Paracchini et al., 2017).

6. CONCLUSION

In contrast to traditional or cryptographic signatures, DNA signatures were not invented in the context of intended intrusions. For practical reasons, the functionality of biologic signatures mostly evolved around balancing sensitivity and specificity. First enhancements to also apprehend some forms of (intended) manipulations were anticipated in Levine (2004) and have become the basis for event-specific GMO detection methods for decades.

The basic idea developed in Levine (2004) incorporates critical components of what is to be expected from a "real" signature. Indeed, the random uptake mechanism of transgenes by *Agrobacterium* creates many gene uptake events of the same transgene into different locations in the host genome. A uniquely identifying event can chosen by selecting both the transgene as well as the accompanying flanking DNA in the host genome. Importantly, due to this mechanism, signatures cannot be reproduced, hence not stolen. It appears that such DNA signatures are at least as reliable as their traditional counterparts. Unfortunately, in the area of modern gene editors, such a signature function alone is not enough. Several risk scenarios have been identified whereby attackers can misuse such signatures via new forms of counterfeiting attacks.

Traditional counterfeiting and blending of high-end products with cheaper material has become a serious problem all around the world. Counterfeit goods have infiltrated most industries from textiles to microchips and pharmaceuticals. And also GMOs! As explained by Berrada et al. (2017), the problem with imitation counterfeits "is that they not only hurt the name of the original and the economy, but because these products are not coming from reliable sources, their quality and efficacy could be compromised."

Counterfeiting by misusing the prevailing DNA signature format may serve several intents beyond mere cost saving. For instance, the recent B2 contamination event in Europe could have been—or analogously, could be—the result of counterfeiting attacks, whereby the producer is falsely blamed for generating unauthorized GMOs. More serious forms of genetic alterations can be envisioned than the mere overproduction of a compound or trait (that may even be authorized in a different jurisdiction). Counterfeiting may also be based on a malicious intent—to not produce cheaper, but hazardous materials or products, as suggested in Mueller (2019). As long as the alterations are outside the signature sequence that is being used to identify the authenticity of the GMO, such forms of intrusions may analogously remain undetected and give evil-doers a way to circulate GMO weapons masquerading as market GMOs.

These new forms of attacks mandate a re-conceptualization of how DNA signatures need to be realized. Herein, several general recommendations have been made that are based on lessons learned from cryptography, overlaid with practical issues that attackers are facing. Based on these, a specific method is suggested that is able to mitigate the misuse of DNA signatures and the distribution of counterfeits.

DNA-signatures enhanced by ZK-based proofs may be extended to different GM organisms or agents. Of special interest here may be emerging pharmaceutical or medical applications, including medicinal products, gene therapy for biological pacemakers (Farraha et al., 2018) and for the nervous system (Bowers et al., 2011). The full potential of supporting cyberbiosecurity risks via cryptographic and cybersecurity means still remains to be fully fleshed out (see also Diggans and Leproust, 2019; Richardson et al., 2019). DNA signatures are just one example where such interdisciplinary insights can lead to effective and safe security solutions.

AUTHOR CONTRIBUTIONS

The author confirms being the sole contributor of this work and has approved it for publication.

ACKNOWLEDGMENTS

I am grateful to Don Roth who supported my interest in the role of biological mediums as information carriers and provided funding for my (second) dissertation (2011–2014). I would also like to thank the reviewers for their careful reading of the manuscript and the many and helpful suggestions they made.

REFERENCES

Allen, J. E., Gardner, S. N., and Slezak, T. R. (2008). DNA signatures for detecting genetic engineering in bacteria. *Genome Biol.* 9:R56. doi: 10.1186/gb-2008-9-3-r56

Arulandhu, A. J., van Dijk, J. P., Dobnik, D., Holst-Jensen, A., Shi, J., Zel, J., et al. (2016). DNA enrichment approaches to identify unauthorized genetically modified organisms (gmos). *Anal. Bioanal. Chem.* 408, 4575–4593. doi: 10.1007/s00216-016-9513-0

Ateniese, G. (2004). Verifiable encryption of digital signatures and applications. *ACM Trans. Inform. Syst. Sec.* 7, 1–20. doi: 10.1145/984334.984335

Barbau-Piednoir, E., Keersmaecker, S. C. J. D., Wuyts, V., Gau, C., Pirovano, W., Costessi, A., et al. (2015). Genome sequence of EU-unauthorized genetically modified bacillus subtilis strain 2014-3557 overproducing riboflavin, isolated from a vitamin B2 80% feed additive. *Genome Announc.* 3:e00214-15. doi: 10.1128/genomeA.00214-15

Berrada, A., Liang, M., Jung, L., and Jensen, K. (2017). *Alkaline Activation for Immobilization of DNA Taggants*. US9790538B2. Available online at: https://patents.google.com/patent/US9790538B2/en

Bowers, W. J., Breakefield, X. O., and Sena-Esteves, M. (2011). Genetic therapy for the nervous system. *Hum. Mol. Genet.* 20, R28–R41. doi: 10.1093/hmg/ddr110

Camenisch, J., and Michels, M. (2000). “Confirmer signature schemes secure against adaptive adversaries,” in *Advances in Cryptology-EUROCRYPT 2000* (Berlin; Heidelberg: Springer), 243–258.

Chaum, D. (1995). “Designated confirmer signatures,” in *Advances in Cryptology-EUROCRYPT’94* (Berlin; Heidelberg: Springer), 86–91.

Chaum, D., and Van Antwerpen, H. (1990). “Undeniable signatures,” in *Advances in Cryptology-CRYPTO’89 Proceedings* (New York, NY: Springer), 212–216.

DiEuliis, D., and Giordano, J. (2017). Why gene editors like CRISPR/Cas may be a game-changer for neuroweapons. *Health Sec.* 15, 296–302. doi: 10.1089/hs.2016.0120

Diggans, J., and Leproust, E. (2019). Next steps for access to safe, secure DNA synthesis. *Front. Bioeng. Biotechn.* 7:86. doi: 10.3389/fbioe.2019.00086

Dunlap, G., and Pauwels, E. (2017). *The Intelligent and Connected Bio-labs of the Future*. Available online at: https://www.wilsoncenter.org/sites/default/files/dunlap_pauwels_intelligent_connected_biolabs_of_future.pdf

EFSA (2011). Panel on genetically modified organisms (GMO). Scientific opinion on guidance on the risk assessment of genetically modified microorganisms and their products intended for food and feed use. *EFSA J.* 9:2193. doi: 10.2903/j.efsa.2011.2193

El Aimani, L. (2009). “On generic constructions of designated confirmer signatures,” in *10th International Conference on Cryptology in India Progress in Cryptology-INDOCRYPT 2009* (Berlin; Heidelberg: Springer), 343–362.

Farraha, M., Kumar, S., Chong, J., Cho, H., and Kizana, E. (2018). Gene therapy approaches to biological pacemakers. *J. Cardiovasc. Dev. Dis.* 5:50. doi: 10.3390/jcdd5040050

Federhen, S., Rossello-Mora, R., Klenk, H.-P., Tindall, B. J., Konstantinidis, K. T., Whitman, W. B., et al. (2016). Meeting report: genbank microbial genomic taxonomy workshop (12–13 may, 2015). *Stand Genomic Sci.* 11:15. doi: 10.1186/s40793-016-0134-1

Frazar, S. L., Hund, G. E., Bonheyo, G. T., Diggans, J., Bartholomew, R. A., Gehrig, L., et al. (2017). Defining the synthetic biology supply chain. *Health Sec.* 15, 392–400. doi: 10.1089/hs.2016.0083

Galbraith, S. D., and Mao, W. (2003). “Invisibility and anonymity of undeniable and confirmer signatures,” in *Cryptographers’ Track at the RSA Conference Topics in Cryptology - CT-RSA 2003* (Berlin; Heidelberg: Springer, LNCS 2612), 80–97.

Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., et al. (2010). Creation of a bacterial cell controlled by a chemically synthesized genome. *Science* 329, 52–56. doi: 10.1126/science.1190719

Golan, E. H., Krissoff, B., Kuchler, F., Calvin, L., Nelson, K. E., and Price, G. K. (2004). *Traceability in the US Food Supply: Economic Theory and Industry Studies*. Technical report. United States Department of Agriculture, Washington, DC. Available online at: https://ageconsearch.umn.edu/record/33939/

Grohmann, L., Keilwagen, J., Duensing, N., Dagand, E., Hartung, F., Wilhelm, R., et al. (2019). Detection and identification of genome editing in plants: challenges and opportunities. *Front. Plant Sci.* 10:236. doi: 10.3389/fpls.2019.00236

Hermann, D., and Schurter, W. (1995). *PCR-Analysis of Riboflavin Samples From Fermentation: Absence of Production Strain Specific DNA*. Unpublished report. WHO, F. Hoffmann La Roche Ltd, Basel.

Holst-Jensen, A., Bertheau, Y., de Loose, M., Grohmann, L., Hamels, S., Hougs, L., et al. (2012). Detecting un-authorized genetically modified organisms (GMOs) and derived materials. *Biotechn. Adv.* 30, 1318–1335. doi: 10.1016/j.biotechadv.2012.01.024

Levine, E. (2004). *Corn Event mon810 and Compositions and Methods for Detection Thereof*. US6713259. Monsanto Technology LLC, St. Louis, MO. Available online at: https://patents.google.com/patent/US6713259

MacKay, D. J. (2003). *Information Theory, Inference and Learning Algorithms*. Cambridge: Cambridge University Press.

Menezes, A. J., Van Oorschot, P. C., and Vanstone, S. A. (1996). *Handbook of Applied Cryptography*. Boca Raton, FL: CRC Press.

Minogue, T. D., Koehler, J. W., Stefan, C. P., and Conrad, T. A. (2019). Next-generation sequencing for biodefense: Biothreat detection, forensics, and the clinic. *Clin. Chem.* 65, 383–392. doi: 10.1373/clinchem.2016.266536

Mueller, S. (2014). *The DNA Code: Implications for Efficiency and Security*. Dissertation, Biomedical Sciences Graduate Ph.D. Program. Laramie, WY: University of Wyoming.

Mueller, S. (2019). Are market gm plants an unrecognized platform for bioterrorism and biocrime? *Front. Bioeng. Biotechn.* 7:121. doi: 10.3389/fbioe.2019.00121

Mueller, S., Jafari, F., and Roth, D. (2015). Improving dependability and precision of data encoding in DNA. *Eur. J. Exp. Biol.* doi: 10.13140/RG.2.1.1215.8325

Mueller, S., Jafari, F., and Roth, D. (2016). A covert authentication and security solution for GMOs. *BMC Bioinform.* 17:389. doi: 10.1186/s12859-016-1256-6

Murch, R. S. (2015). Bioattribution needs a coherent international approach to improve global biosecurity. *Front. Bioeng. Biotechn.* 3:80. doi: 10.3389/fbioe.2015.00080

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechn.* 6:39. doi: 10.3389/fbioe.2018.00039

Paracchini, V., Petrillo, M., Reiting, R., Angers-Loustau, A., Wahler, D., Stolz, A., et al. (2017). Molecular characterization of an unauthorized genetically modified bacillus subtilis production strain identified in a vitamin b 2 feed additive. *Food Chem.* 230, 681–689. doi: 10.1016/j.foodchem.2017.03.042

Peccoud, J., Anderson, J. C., Chandran, D., Densmore, D., Galdzicki, M., Lux, M. W., et al. (2011). Essential information for synthetic DNA sequences. *Nat. Biotechn.* 29:22. doi: 10.1038/nbt0111-22b

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2018). Cyberbiosecurity: From naive trust to risk awareness. *Trends Biotechn.* 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012

Permingeat, H. R., Reggiardo, M. I., and Vallejos, R. H. (2002). Detection and quantification of transgenes in grains by multiplex and real-time PCR. *J. Agric. Food Chem.* 50, 4431–4436. doi: 10.1021/jf020081d

Phillippy, A. M., Mason, J. A., Ayanbule, K., Sommer, D. D., Taviani, E., Huq, A., et al. (2007). Comprehensive DNA signature discovery and validation. *PLoS Comput. Biol.* 3:e98. doi: 10.1371/journal.pcbi.0030098

Richardson, L. C., Connell, N. D., Lewis, S. M., Pauwels, E., and Murch, R. S. (2019). Cyberbiosecurity: a call for cooperation in a new threat landscape. *Front. Bioeng. Biotechnol.* 7:99. doi: 10.3389/fbioe.2019.00099

Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E., and Lippman, Z. B. (2017). Engineering quantitative trait variation for crop improvement by genome editing. *Cell* 171, 470–480. doi: 10.1016/j.cell.2017.08.030

Stinson, D. R. (2005). *Cryptography: Theory and Practice*. New York, NY: Chapman and Hall/CRC.

Wilson, M. L., Cai, Y., Hanlon, R., Taylor, S., Chevreux, B., Setubal, J. C., et al. (2012). Sequence verification of synthetic DNA by assembly of sequencing reads. *Nucleic Acids Res.* 41, e25–e25. doi: 10.1093/nar/gks908

Xia, F. (2013). *Designated Confirmer Signatures: Modelling, Design and Analysis*. Ph.D. thesis. Birmingham: University of Birmingham.

Conflict of Interest Statement: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PERSPECTIVE
published: 04 June 2019
doi: 10.3389/fbioe.2019.00128

Perspectives on Harmful Algal Blooms (HABs) and the Cyberbiosecurity of Freshwater Systems

David G. Schmale III[1]*, *Andrew P. Ault*[2,3], *Walid Saad*[4], *Durelle T. Scott*[5] and *Judy A. Westrick*[6]

[1] School of Plant and Environmental Sciences, Virginia Tech, Blacksburg, VA, United States, [2] Department of Environmental Health Sciences, University of Michigan, Ann Arbor, MI, United States, [3] Department of Chemistry, University of Michigan, Ann Arbor, MI, United States, [4] Bradley Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA, United States, [5] Department of Biological Systems Engineering, Virginia Tech, Blacksburg, VA, United States, [6] Lumigen Instrumentation Center, Department of Chemistry, Wayne State University, Detroit, MI, United States

OPEN ACCESS

Edited by:
Diane DiEuliis,
National Defense University,
United States

Reviewed by:
Maria José Saavedra,
Universidade de Trás os Montes e Alto Douro, Portugal
Deon Van Der Merwe,
GD Animal Health, Netherlands

***Correspondence:**
David G. Schmale III
dschmale@vt.edu

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *22 February 2019*
Accepted: *13 May 2019*
Published: *04 June 2019*

Citation:
Schmale DG III, Ault AP, Saad W, Scott TD and Westwrick AJ (2019) Perspectives on Harmful Algal Blooms (HABs) and the Cyberbiosecurity of Freshwater Systems. Front. Bioeng. Biotechnol. 7:128. doi: 10.3389/fbioe.2019.00128

Harmful Algal Blooms (HABs) have been observed in all 50 states in the U.S., ranging from large freshwater lakes, such as the Great Lakes, to smaller inland lakes, rivers, and reservoirs, as well as marine coastal areas and estuaries. In 2014, a HAB on Lake Erie containing microcystin (a liver toxin) contaminated the municipal water supply in Toledo, Ohio, providing non-potable water to 400,000 people. Studying HABs is complicated as different cyanobacteria produce a range of toxins that impact human health, such as microcystins, saxitoxin, anatoxin-a, and cylindrospermopsin. HABs may be increasing in prevalence with rising temperatures and higher nutrient runoff. Consequently, new tools and technology are needed to rapidly detect, characterize, and respond to HABs that threaten our water security. A framework is needed to understand cyber threats to new and existing technologies that monitor and forecast our water quality. To properly detect, assess, and mitigate security threats on water infrastructure, it is necessary to envision water security from the perspective of a cyber-physical system (CPS). In doing so, we can evaluate risks and research needs for cyber-attacks on HAB-monitoring networks including data injection attacks, automated system hijacking attacks, node forgery attacks, and attacks on learning algorithms. Herein, we provide perspectives on the research needed to understand both the threats posed by HABs and the coupled cyber threats to water security in the context of HABs.

Keywords: harmful algal bloom, cyanobacteria, algae, toxin, water security, cybersecurity, drone

DISTRIBUTION AND TRANSPORT OF HABS IN THE UNITED STATES.

The intensity and frequency of harmful algal blooms (HABs) has increased globally in recent years (Backer et al., 2015). In the U.S., HABs have been observed in variety of freshwater ecosystems including the Great Lakes (**Figure 1**, left), small inland lakes, and rivers. Consequently, new legislation has been developed to protecting the general public from HABs (National Science Technology Council Subcommittee on Ocean Science Technology, 2016). In 2014, an HAB caused by *Microcystis* at the water treatment plant intake for Toledo, Ohio led to

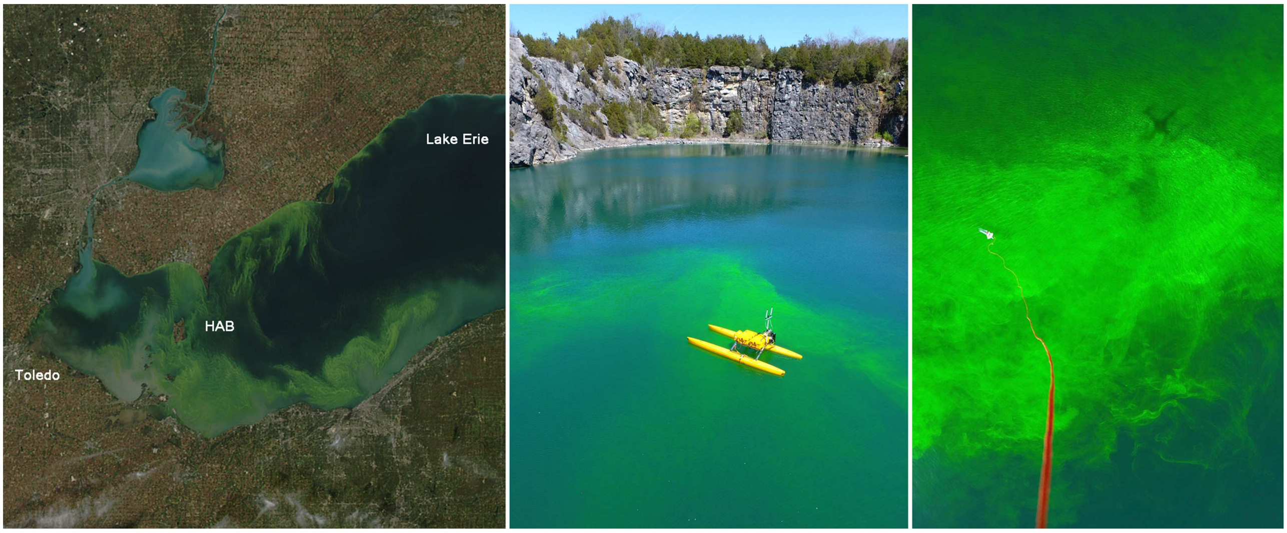


FIGURE 1 | Harmful algal bloom (HAB) in Lake Erie, USA on October 9, 2011 as recorded by Moderate Resolution Imaging Spectroradiometer (MODIS) on the Aqua satellite **(Left)** (courtesy NASA). Technologies with unmanned systems in the water (center panel) and air **(Right)** have the potential to be used to monitor HABs *in situ*. The unmanned systems shown here were tuned to a released fluorescein dye, which has been used a surrogate for HABs (Powers et al., 2018a) (courtesy D. Schmale).

the distribution of non-potable water for multiple days (Steffen et al., 2017). Increasing concerns related to health are not limited to *Microcystis*, as many other genera of cyanobacteria (*Planktothrix*, *Alexandrium*, *Anabaena*, *Cylindrospermopsis, Euglena, etc.*) and associated toxins (anatoxin-a, saxitoxins, cylindrospermopsin, euglenophycin, etc.) have been observed in a range of freshwater systems (Graham et al., 2010; Foss and Aubel, 2015; Loftin et al., 2016; Birbeck et al., 2019). Today, a critical gap still exists on the relationship between human and animal health impact and the range of conditions under which the toxins are produced and the resulting range of toxicities (~50 to >5,000 μg kg^{-1}) (Chorus and Bartram, 1999). In 2016, the United States Environmental Protection Agency (EPA) released a draft health advisory for recreational exposure of 4 μg L^{-1} for microcystins and 8 μg L^{-1} for cylindrospermopsin, which are lower than the 20 μg L^{-1} limit recommended by the World Health Organization (WHO) (Environmental Protection Agency, 2016). The lower levels in the draft advisory will result in more frequent exceedance conditions, and highlight that lower concentrations can be harmful. The cyanobacteria and toxins produced by HABs vary considerably within and between freshwater ecosystems, making predictions of HAB formation and toxin production challenging and beyond the capabilities of current models.

Knowledge of the transport of HAB toxins between blooms and the point of human exposure is not well-understood and is crucial for protecting the public. Extracellular material from freshwater cyanobacteria blooms has been observed in the water and in the atmosphere after aerosolization at locations far beyond the edges of HABs (Wood and Dietrich, 2011). For example, extracellular toxins from blooms (e.g., microcystins) have been observed downstream of HABs due to upstream treatment with algicide (Graham et al., 2012). Aerosolized biogenic organic material from freshwater blooms has also been observed in the atmosphere above HABs and transported at least 30 km inland (May et al., 2017, 2018). Thus, it is not only critical to understand when blooms form and produce toxins, but also to understand transport of algal toxins in water and air beyond the bloom.

To improve the security of our freshwater resources it is critical to improve our understanding of HABs and toxin exposures to the point of making predictions that can be used to guide policy and protect public health. This will require transdisciplinary research at the nexus of ecology, atmospheric science, chemistry, and engineering to understand the threats, manage risks, and develop capabilities for reducing exposure to HAB toxins. Challenges to understanding the distribution and transport of HABs include the changing climate, where warmer temperatures facilitate greater HAB formation and changes in nutrient loadings (nitrogen and phosphorus) (Ho and Michalak, 2017; Del Giudice et al., 2018) that are likely to only improve conditions for more intense HABs in many locations across the U.S. (Kosten et al., 2012; Stocker et al., 2013). Remote sensing can provide valuable information regarding the density, extent, and potential impact of blooms that are known to be harmful (Ho and Michalak, 2015; Ho et al., 2017). However, these methods cannot actually determine if an algal bloom is harmful, since they do not monitor the toxins being produced and released. Addressing the challenging of understanding toxin temporal and spatial distributions will require utilizing innovative sampling and direct toxin measurements to improve detection and monitoring capabilities. Such sampling methods must consider the dynamics of the bloom and toxins in relation to the spatial and temporal sampling and testing capacity, to assess the viability of the technology to achieve the goals. For example, unmanned systems (drones) in the air (Benson et al., 2019) and water (Powers et al., 2018a,b) have the potential to be used to detect, track, and sample

HABs from the atmosphere and freshwater systems (**Figure 1**, right). Analytical chemistry approaches can be used to quantify specific HAB toxins from these samples, and to determine chemical signatures to predict toxin production. Ultimately, predictive understanding of freshwater HAB toxin production and transport is essential to improve the security of freshwater systems and protect public health in the many regions with increasingly intense HABs.

DETECTION AND CHARACTERIZATION OF TOXINS ASSOCIATED WITH HABS

Although there are many classes of cyanotoxins (Janssen, 2019), microcystin (MC), nodularin (NOD), saxitoxin (SXT), anatoxin-a (ANA), and cylindrospermopsin (CYL) are monitored frequently and have been reported in recreational water (Chorus et al., 2000; Graham et al., 2010), drinking water sources (Otten and Paerl, 2015; Steffen et al., 2017), and potable water that was believed to have been treated (Steffen et al., 2017; Davis et al., 2019). Most researchers and water practitioners use commercial semiquantitative cyanotoxin detection technologies because these platforms are inexpensive and need less technical skill (e.g., Humpage et al., 2012; Aranda-Rodriguez et al., 2015). Important limitations of testing technologies within the context of HAB-assessment include throughput capacity, lag times between sampling and results, and the infrastructure/equipment/people needed for implementation.

Commercial semi-quantitative technologies include cyanotoxin class specific enzyme inhibition assays, enzyme-linked immunoassays (ELISAs), and strip tests. Although these technologies are relatively inexpensive, simple and rapid, these assays have a narrow standard dynamic range (0.15 to 5 ppb cyanotoxin) and are not as selective as mass spectrometry (Westrick and Szlag, 2018). Only one enzyme inhibition assay is commercially available, the MC and NOD protein phosphatase inhibition assay (PPIA). Since the MC or NOD Adda amino acid irreversibly binds to the protein phosphatase active site, scientists often refer to the PPIA as a toxicity assay (Carmichael and An, 1999). Limited publications have reported using the commercial protein phosphatase inhibition assay (PPIA); more peer review literature is needed to determine its efficacy (Gaget et al., 2017). Cyanotoxin ELISAs and strip tests are based on polyclonal antibody technology (Humpage et al., 2012; Aranda-Rodriguez et al., 2015). The primary concern with ELISA is that each class of cyanotoxin has several variants or congeners with different toxicities. Usually the antibody is raised against a part of the congener, and does not bind equally to all congeners presenting cross-reactivity effects (Fischer et al., 2001). Antibodies and antibody assays are commercially available for MC, SXT, ANA, and CYL (Westrick and Szlag, 2018). US EPA and several states have required the use of an ELISA where the antibodies are raised against Adda-haptens, providing cross-reactivity to all MC congeners (Fischer et al., 2001). Since all kits use microcystin-specific-LR (MCLR) calibration curves the ELISA units are often called "total" MCLR equivalents (total MCLReq). However, the degree of binding to the antibody does not have a relationship to toxicity (Metcalf et al., 2000), as recent publications suggest that biodegradation and oxidation products interfere with the antibody and provide inaccurate quantitation (Guo et al., 2017; He et al., 2017; Thees et al., 2018).

Liquid chromatography (LC) with various detectors, such as photodiode array (PDA), fluorometer (FL), and tandem mass spectrometer (MS/MS) has been used to quantitate cyanotoxins (Meriluoto et al., 2017). LC/PDA provides detection limits around 0.2 ppb for all MCs because the Adda amino acid produces an ultraviolet spectrum with a maximum absorbance at 238 nm with an extinction coefficient around 39,000 L mol^{-1} cm^{-1} (ISO, 2005). Saxitoxins are derivatized into a fluorescence compound and analyzed by LC/FL. LC/MS/MS methods have been developed for SXT (Onodera et al., 1997), MC (Triantis et al., 2016; Turner et al., 2018), CYL (Guzmán-Guillén et al., 2012), and ANA (Sanchez et al., 2014; Wood et al., 2017). Several MS/MS methods can identify and quantify multiple classes of toxins (Oehrle et al., 2010) with a an expanded dynamic range of about 0.1 to 1,000 ppb beyond the commercial techniques and without the need for solid phase extraction to concentrate the sample. Advancement of a one-step technology, LC/MS/MS with online concentration, enables quantification in under 10 min and provides a dynamic range to 0.005 ppb to 1 ppb (Flores and Caixach, 2017; Birbeck et al., 2019). However, many cyanotoxin quantitation challenges remain. Key issues include known and unknown structural diversity in each cyanotoxin class and between cyanotoxin classes, as well as the lack of reference materials. In order to provide HAB cyber-monitoring, new real-time and passive analytical micro technologies need to be developed and incorporated into unmanned systems for monitoring HABS, such as drones, boats, and buoys (as discussed in the previous section). These unmanned systems can co-locate the sensors with the sampling mechanism, reducing lag times associated with sample transport to a testing facility and improving the applicability of data in near real-time risk assessments.

TRENDS IN HABS RELATED TO HIGHER TEMPERATURES AND NUTRIENT RUNOFF

Balancing food production for the world's growing population while maintaining our water resources is one of society's larger challenges (Foley et al., 2011). Across the globe, our inland freshwater and coastal zones are experiencing widespread eutrophication (Diaz and Rosenberg, 2008), resulting in declining oxygen concentrations (Breitburg et al., 2018), and harmful algal blooms (Anderson et al., 2002). While excess fertilizer use and manure waste has been recognized as an issue for freshwater systems for over two decades (Carpenter et al., 1998), nitrogen and phosphorus runoff into streams and rivers continues to increase. In addition to increasing export, a recent global analysis of riverine nutrient export found a larger proportion of inorganic nitrogen, and phosphorus (Vilmin et al., 2018), altering nutrient ratios relevant for phytoplankton and algal communities.

Nitrogen and phosphorus sources are attributed to fertilizers, animal waste, atmospheric deposition, and municipal sewage. For nitrogen, the Haber-Bosch process has fundamentally altered the nitrogen cycle (Galloway et al., 2004) by providing a mechanism to convert N_2 to reduced nitrogen for use in fertilizers. Worldwide, fertilizer production continues to increase, largely in the form of urea (Glibert et al., 2006). The primary source of phosphorus fertilizer is from mining. Fossil-fuel combustion and land-conversion over the last century have also provided a source of reactive nitrogen to the atmosphere, which in turn is transported through the atmosphere beyond the emission source leaving no landscape untouched, even the most pristine (Elser et al., 2009).

While fertilizers are required for plant growth, agricultural landscapes are connected to freshwater systems both during dry and wet periods. During storm events and snowmelt, water is transported either across the ground surface or infiltrates through the soil matrix into groundwater and can mobilize nutrients. In regions with poorly draining soils, tile drainage is also used to efficiently remove excess water from a field and can lead to removal of >75% of the water (Van Esbroeck et al., 2016). While these systems maximize food production (Chowdhury et al., 2017), they also short circuit riparian zones and reduce nutrient retention. The dominant forms of nitrogen are generally dissolved, and will move in both surface runoff and into groundwater. Soil erosion is largely thought to be the primary source of phosphorus (Vilmin et al., 2018), although soluble phosphorus is also mobilized and transported through tile drainage (Smith et al., 2015).

Beyond the challenge of excess fertilizer use, the other challenge is our changing climate which may result in higher magnitude storms followed by droughts. While best management practices (BMP) can be adopted, many BMPs may result in nutrient pulses to freshwater systems, which has been shown to increase HAB development (Spatharis et al., 2007). Municipal waste provides a continual nutrient source directly into freshwaters even in times of drought (Mosley, 2015; Vilmin et al., 2018). Thus, while efforts to reduce excess nutrient applications and use of BMPs are needed, our changing climate makes effective solutions challenging (Scavia et al., 2014).

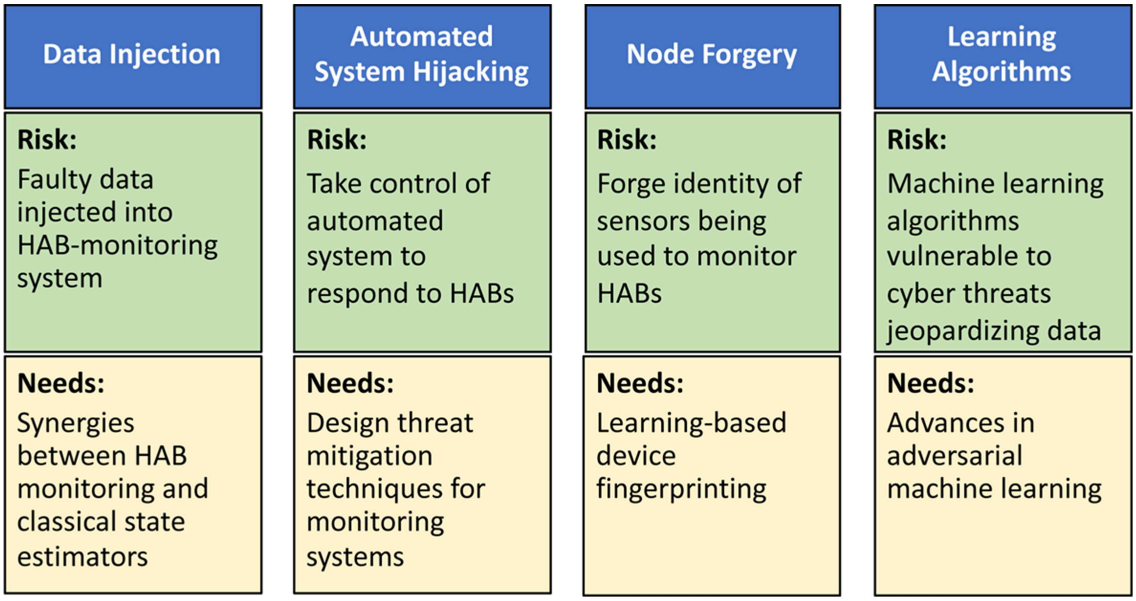
Warmer temperatures are predicted to increase HAB formation and toxin release (O'neil et al., 2012). Two primary mechanisms play a role: increased growth rates and greater stratification and water column stability. Lake experiments have observed higher cyanobacteria growth rates in response to higher temperatures (Liu et al., 2011). In one experimental study, increased water temperature resulted in significant increases in microcystis growth rates (Davis et al., 2019). This research was supported by another study that reported higher growth rates in response to temperature and phosphorus availability (Duan et al., 2009). These HAB responses to warmer water result in greater microcystin toxin releases up to a temperature threshold (Walls et al., 2018). The second mechanism for increased HAB are from enhanced stratification (Joehnk et al., 2008; Rabalais et al., 2009; Paerl et al., 2011). With climate variability, longer, and hotter summers with more frequent droughts are likely to increase water column stability (Mosley, 2015), changing the competition dynamics. Cyanobacteria are able to migrate vertically, providing a competitive advantage (Huisman et al., 2004; Lürling et al., 2013). Future HAB management will require addressing not only nutrient management and runoff, but also thermal regimes and stratification (Paerl et al., 2011), given the body of evidence suggesting the confounding effects from nutrients, temperature, and thermal stratification. Furthermore, integrating monitoring technology that includes the key factors for HAB formation required for forecasting and subsequent HAB and toxin development requires not only improved approaches for toxin quantification and detection but a holistic, secure cyber-physical system (CPS), as described in the next section.

FRAMEWORKS AND RESEARCH NEEDS FOR WATER CYBER-PHYSICAL SECURITY

HABs pose physical threats on water security. However, to properly detect, assess, and mitigate security threats on water infrastructure, it is imperative to envision water security from the perspective of a cyber-physical system (CPS). Indeed, our national water infrastructure can be seen as a CPS whose physical realm pertains to the physical body of water and the devices that directly interact with it and whose cyber realm pertains to the sensors, smart meters, and other networked apparatus that connect the physical system to the Internet.

In the context of HABs, we envision four types of cyber-attacks with physical targets: data injection, automated system hijacking, node forgery, and learning algorithms (**Figure 2**). In data injection attacks, adversaries can inject faulty data, through the HAB monitoring system, to mislead it into underestimating HAB levels. If undetected, such faulty data can potentially lead to a physical catastrophe on the monitored body of water (Moyer et al., 2009). A robust water cybersecurity program is essential to protect public health and prevent service disruptions (Panguluri et al., 2017). Lessons learned from power systems (Sanjab and Saad, 2016) show that such attacks can be done stealthily without being detected by standard state estimators. Meanwhile, automated system hijacking attacks can be launched to take control of any automated system (e.g., an automated drone or sensors) used to respond to rising HAB levels. By taking control of the automated HAB response system, the adversary can derail the system from its original mission thus once again jeopardizing water security. Moreover, the low-cost, small form factor nature of monitoring sensors renders them highly vulnerable to node forgery attacks where an adversary can forge the identity of monitoring sensors and use those captured sensors to jeopardize the integrity of the HAB monitoring data being collected. Finally, the need for data analytics in HAB monitoring will involve machine learning algorithms whose operation will be vulnerable to cyber threats that can jeopardize their input and output data.

Clearly, it is imperative to develop new techniques to mitigate the aforementioned cyber-physical security threats (**Figure 2**). To this end, as shown in Sanjab and Saad (2016) and Ferdowsi et al. (2017), one can leverage tools from game theory to understand how defenders and attackers interact over a water CPS and, therefore, identify potential vulnerabilities and optimal



Data Injection	Automated System Hijacking	Node Forgery	Learning Algorithms
Risk: Faulty data injected into HAB-monitoring system	**Risk:** Take control of automated system to respond to HABs	**Risk:** Forge identity of sensors being used to monitor HABs	**Risk:** Machine learning algorithms vulnerable to cyber threats jeopardizing data
Needs: Synergies between HAB monitoring and classical state estimators	**Needs:** Design threat mitigation techniques for monitoring systems	**Needs:** Learning-based device fingerprinting	**Needs:** Advances in adversarial machine learning

FIGURE 2 | Risks and research needs for cyber-attacks on HAB-monitoring systems including data injection attacks, automated system hijacking attacks, node forgery attacks, and attacks on learning algorithms.

cyber-physical defense strategies. This theoretical analysis can then be used to develop various software/hardware solutions, such as improved estimation algorithms and robust control mechanisms (Ferdowsi et al., 2018), to mitigate data injection and autonomous system hijack threats. To deal with node forgery attacks, the use of learning-based device fingerprinting (Ferdowsi and Saad, 2018) can be particularly apropos. In addition, advances in adversarial machine learning (McDaniel et al., 2016) can provide tools to develop robust data analytics in water systems. However, additional research is needed to tailor existing solutions to the unique properties of water systems.

In terms of data injection attacks, research is needed to understand the synergies and distinctions between HAB monitoring systems and classical state estimators, such as those used in power systems. This understanding is necessary to devise HAB-specific solutions for mitigating data injection attacks. Water infrastructure will not only depend on the cyber infrastructure, but it will also be interconnected with other infrastructure in a city. This interdependence and its impact on CPS security must be identified and analyzed. In terms of automated system hijacking attacks, it is necessary to first introduce HAB-centric automated systems that can effectively help in monitoring and treating HAB-affected bodies of waters. Once such systems are in place, one can better design threat mitigation techniques that are tailored to those systems by leveraging on lessons learned from other fields, such as autonomous vehicles (Ferdowsi et al., 2018).

In summary, new analytical tools that can build on existing techniques such as graph theory and game theory, are needed to devise realistic models for water systems, in general, and HAB-centric water monitoring systems in particular. Such models must capture the physical dynamics of the system as well as the cyber-interconnections. By devising such models, potential vulnerabilities can be better identified and new strategies for securing the system can be devised. Research developing a convergent paradigm at the nexus of engineering, ecology, and chemistry is needed to understand threats, manage risks, and develop capabilities for water cyberbiosecurity.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

This manuscript was supported in part by a cybersecurity grant from the College of Agriculture and Life Sciences at Virginia Tech and a grant to JW number P30 ES020957.

REFERENCES

Anderson, D. M., Glibert, P. M., and Burkholder, J. M. (2002). Harmful algal blooms and eutrophication: nutrient sources, composition, and consequences. *Estuaries* 25, 704–726. doi: 10.1007/BF02804901

Aranda-Rodriguez, R., Jin, Z., Harvie, J., and Cabecinha, A. (2015). Evaluation of three field test kits to detect microcystins from a public health perspective. *Harmful Algae* 42, 34–42. doi: 10.1016/j.hal.2015.01.001

Backer, L. C., Manassaram-Baptiste, D., LePrell, R., and Bolton, B. (2015). Cyanobacteria and algae blooms: review of health and environmental data from the harmful algal bloom-related illness surveillance system (HABISS) 2007–2011. *Toxins* 7, 1048–1064. doi: 10.3390/toxins7041048

Benson, J., Hanlon, R., Seifried, T. M., Baloh, P., Powers, C. W., Grothe, H., et al. (2019). Microorganisms collected from the surface of freshwater lakes using a drone water sampling system (DOWSE). *Water* 11:157. doi: 10.3390/w11010157

Birbeck, J., Westrick, J., O'Neill, G., Spies, B., and Szlag, D. (2019). Comparative analysis of microcystin prevalence in michigan lakes by online concentration LC/MS/MS and ELISA. *Toxins* 11:13. doi: 10.3390/toxins11010013

Breitburg, D., Levin, L. A., Oschlies, A., Grégoire, M., Chavez, F. P., Conley, D. J., et al. (2018). Declining oxygen in the global ocean and coastal waters. *Science* 359:eaam7240. doi: 10.1126/science.aam7240

Carmichael, W. W., and An, J. (1999). Using an enzyme linked immunosorbent assay (ELISA) and a protein phosphatase inhibition assay (PPIA) for the detection of microcystins and nodularins. *Nat. Toxins* 7, 377–385. doi: 10.1002/1522-7189(199911/12)7:6&lt;377::AID-NT80&gt;3.0.CO;2-8

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., and Smith, V. H. (1998). Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecol. Appl.* 8, 559–568. doi: 10.1890/1051-0761(1998)008[0559:NPOSWW]2.0.CO;2

Chorus, I., and Bartram, J. (1999). *Toxic Cyanobacteria in Water: A Guide to Their Public Health Consequences, Monitoring and Management*. London: CRC Press. doi: 10.4324/9780203478073

Chorus, I., Falconer, I. R., Salas, H. J., and Bartram, J. (2000). Health risks caused by freshwater cyanobacteria in recreational waters. *J. Toxicol. Environ. Health B Crit. Rev.* 3, 323–347. doi: 10.1080/109374000436364

Chowdhury, R. B., Moore, G. A., Weatherley, A. J., and Arora, M. (2017). Key sustainability challenges for the global phosphorus resource, their implications for global food security, and options for mitigation. *J. Clean. Prod.* 140, 945–963. doi: 10.1016/j.jclepro.2016.07.012

Davis, T. W., Stumpf, R., Bullerjahn, G. S., McKay, R. M. L., Chaffin, J. D., Bridgeman, T. B., et al. (2019). Science meets policy: a framework for determining impairment designation criteria for large waterbodies affected by cyanobacterial harmful algal blooms. *Harmful Algae* 81, 59–64. doi: 10.1016/j.hal.2018.11.016

Del Giudice, D., Zhou, Y., Sinha, E., and Michalak, A. M. (2018). Long-term phosphorus loading and springtime temperatures explain interannual variability of hypoxia in a large temperate lake. *Environ. Sci. Technol.* 52, 2046–2054. doi: 10.1021/acs.est.7b04730

Diaz, R. J., and Rosenberg, R. (2008). Spreading dead zones and consequences for marine ecosystems. *Science* 321, 926–929. doi: 10.1126/science.1156401

Duan, H., Ma, R., Xu, X., Kong, F., Zhang, S., Kong, W., et al. (2009). Two-decade reconstruction of algal blooms in China's Lake Taihu. *Environ. Sci. Technol.* 43, 3522–3528. doi: 10.1021/es8031852

Elser, J. J., Andersen, T., Baron, J. S., Bergström, A.-K., Jansson, M., Kyle, M., et al. (2009). Shifts in lake N: P stoichiometry and nutrient limitation driven by atmospheric nitrogen deposition. *science* 326, 835–837. doi: 10.1126/science.1176199

Environmental Protection Agency (2016). *Draft Human Health Recreational Ambient Water Quality Criteria and/or Swimming Advisories for Microcystins and Cylindrospermopsin*. Available online at: https://www.federalregister.gov/documents/2016/12/19/2016-30464/request-for-scientific-views-draft-human-health-recreational-ambient-water-quality-criteria-andor (accessed May 23, 2019).

Ferdowsi, A., Ali, S., Saad, W., and Mandayam, N. B. (2018). Cyber-physical security and safety of autonomous connected vehicles: optimal control meets multi-armed bandit learning. *ArXiv Prepr. ArXiv*181205298.

Ferdowsi, A., and Saad, W. (2018). Deep learning for signal authentication and security in massive Internet of Things systems. *ArXiv Prepr. ArXiv*180300916.

Ferdowsi, A., Saad, W., and Mandayam, N. B. (2017). Colonel blotto game for secure state estimation in interdependent critical infrastructure. *ArXiv Prepr. ArXiv*170909768.

Fischer, W. J., Garthwaite, I., Miles, C. O., Ross, K. M., Aggen, J. B., Chamberlin, A. R., et al. (2001). Congener-independent immunoassay for microcystins and nodularins. *Environ. Sci. Technol.* 35, 4849–4856. doi: 10.1021/es011182f

Flores, C., and Caixach, J. (2017). "Analysis of microcystins by online solid phase extraction–liquid chromatography tandem mass spectrometry," in *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis*, eds. J. Meriluoto, L. Spoof, and G. A. Codd (Chichester, UK: John Wiley and Sons), 362–371. doi: 10.1002/9781119068761.ch41

Foley, J. A., Ramankutty, N., Brauman, K. A., Cassidy, E. S., Gerber, J. S., Johnston, M., et al. (2011). Solutions for a cultivated planet. *Nature* 478, 337–334. doi: 10.1038/nature10452

Foss, A. J., and Aubel, M. T. (2015). Using the MMPB technique to confirm microcystin concentrations in water measured by ELISA and HPLC (UV, MS, MS/MS). *Toxicon* 104, 91–101. doi: 10.1016/j.toxicon.2015.07.332

Gaget, V., Lau, M., Sendall, B., Froscio, S., and Humpage, A. R. (2017). Cyanotoxins: which detection technique for an optimum risk assessment? *Water Res.* 118, 227–238. doi: 10.1016/j.watres.2017.04.025

Galloway, J. N., Dentener, F. J., Capone, D. G., Boyer, E. W., Howarth, R. W., Seitzinger, S. P., et al. (2004). Nitrogen cycles: past, present, and future. *Biogeochemistry* 70, 153–226. doi: 10.1007/s10533-004-0370-0

Glibert, P. M., Harrison, J., Heil, C., and Seitzinger, S. (2006). Escalating worldwide use of urea–a global change contributing to coastal eutrophication. *Biogeochemistry* 77, 441–463. doi: 10.1007/s10533-005-3070-5

Graham, J. L., Loftin, K. A., Meyer, M. T., and Ziegler, A. C. (2010). Cyanotoxin mixtures and taste-and-odor compounds in cyanobacterial blooms from the Midwestern United States. *Environ. Sci. Technol.* 44, 7361–7368. doi: 10.1021/es1008938

Graham, J. L., Ziegler, A. C., Loving, B. L., and Loftin, K. A. (2012). *Fate and Transport of Cyanobacteria and Associated Toxins and Taste-and-Odor Compounds From Upstream Reservoir Releases in the Kansas River, Kansas, September and October 2011*. US Department of the Interior, US Geological Survey. doi: 10.3133/sir20125129

Guo, Y. C., Lee, A. K., Yates, R. S., Liang, S., and Rochelle, P. A. (2017). Analysis of microcystins in drinking water by ELISA and LC/MS/MS. *J. Am. Water Works Assoc.* 109, 13–25. doi: 10.5942/jawwa.2017.109.0027

Guzmán-Guillén, R., Prieto, A. I., González, A. G., Soria-Díaz, M. E., and Cameán, A. M. (2012). Cylindrospermopsin determination in water by LC-MS/MS: optimization and validation of the method and application to real samples. *Environ. Toxicol. Chem.* 31, 2233–2238. doi: 10.1002/etc.1954

He, X., Stanford, B. D., Adams, C., Rosenfeldt, E. J., and Wert, E. C. (2017). Varied influence of microcystin structural difference on ELISA cross-reactivity and chlorination efficiency of congener mixtures. *Water Res.* 126, 515–523. doi: 10.1016/j.watres.2017.09.037

Ho, J. C., and Michalak, A. M. (2015). Challenges in tracking harmful algal blooms: a synthesis of evidence from Lake Erie. *J. Gt. Lakes Res.* 41, 317–325. doi: 10.1016/j.jglr.2015.01.001

Ho, J. C., and Michalak, A. M. (2017). Phytoplankton blooms in Lake Erie impacted by both long-term and springtime phosphorus loading. *J. Gt. Lakes Res.* 43, 221–228. doi: 10.1016/j.jglr.2017.04.001

Ho, J. C., Stumpf, R. P., Bridgeman, T. B., and Michalak, A. M. (2017). Using Landsat to extend the historical record of lacustrine phytoplankton blooms: a lake erie case study. *Remote Sens. Environ.* 191, 273–285. doi: 10.1016/j.rse.2016.12.013

Huisman, J., Sharples, J., Stroom, J. M., Visser, P. M., Kardinaal, W. E. A., Verspagen, J. M., et al. (2004). Changes in turbulent mixing shift competition for light between phytoplankton species. *Ecology* 85, 2960–2970. doi: 10.1890/03-0763

Humpage, A. R., Froscio, S. M., Lau, H.-M., Murphy, D., and Blackbeard, J. (2012). Evaluation of the abraxis strip test for microcystins^{tm} for use with wastewater effluent and reservoir water. *Water Res.* 46, 1556–1565. doi: 10.1016/j.watres.2011.12.015

ISO (2005). *Water Quality–Determination of Microcystins–Method Using Solid Phase Extraction (SPE) and High Performance Liquid Chromatography (HPLC) with Ultraviolet (UV) Detection*.

Janssen, E. M.-L. (2019). Cyanobacterial peptides beyond microcystins–A review on co-occurrence, toxicity, and challenges for risk assessment. *Water Res.*151, 488–499. doi: 10.1016/j.watres.2018.12.048

Joehnk, K. D., Huisman, J. E. F., Sharples, J., Sommeijer, B. E. N., Visser, P. M., and Stroom, J. M. (2008). Summer heatwaves promote blooms of harmful cyanobacteria. *Glob. Change Biol.* 14, 495–512. doi: 10.1111/j.1365-2486.2007.01510.x

Kosten, S., Huszar, V. L., Bécares, E., Costa, L. S., Van Donk, E., Hansson, L.-A., et al. (2012). Warmer climates boost cyanobacterial dominance in shallow lakes. *Glob. Change Biol.* 18, 118–126. doi: 10.1111/j.1365-2486.2011.02488.x

Liu, X., Lu, X., and Chen, Y. (2011). The effects of temperature and nutrient ratios on Microcystis blooms in Lake Taihu, China: an 11-year investigation. *Harmful Algae* 10, 337–343. doi: 10.1016/j.hal.2010.12.002

Loftin, K. A., Graham, J. L., Hilborn, E. D., Lehmann, S. C., Meyer, M. T., Dietze, J. E., et al. (2016). Cyanotoxins in inland lakes of the United States: occurrence and potential recreational health risks in the EPA National Lakes Assessment 2007. *Harmful Algae* 56, 77–90. doi: 10.1016/j.hal.2016.04.001

Lürling, M., Eshetu, F., Faassen, E. J., Kosten, S., and Huszar, V. L. (2013). Comparison of cyanobacterial and green algal growth rates at different temperatures. *Freshw. Biol.* 58, 552–559. doi: 10.1111/j.1365-2427.2012.02866.x

May, N. W., Gunsch, M. J., Olson, N., Bondy, A. L., Kirpes, R. M., Bertman, S., et al. (2018). Unexpected contributions of sea spray and lake spray aerosol to inland particulate matter. *Environ. Sci. Technol. Lett.* 5, 405–512. doi: 10.1021/acs.estlett.8b00254

May, N. W., Olson, N. E., Panas, M., Axson, J. L., Tirella, P. S., Kirpes, R. M., et al. (2017). Aerosol emissions from Great Lakes harmful algal blooms. *Environ. Sci. Technol.* 52, 397–405. doi: 10.1021/acs.est.7b03609

McDaniel, P., Papernot, N., and Celik, Z. B. (2016). Machine learning in adversarial settings. *IEEE Secur. Priv.* 14, 68–72. doi: 10.1109/MSP.2016.51

Meriluoto, J., Spoof, L., and Codd, G. A. (2017). *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis*. Chichester, UK: John Wiley and Sons. doi: 10.1002/9781119068761

Metcalf, J. S., Beattie, K. A., Pflugmacher, S., and Codd, G. A. (2000). Immuno-crossreactivity and toxicity assessment of conjugation products of the cyanobacterial toxin, microcystin-LR. *FEMS Microbiol. Lett.* 189, 155–158. doi: 10.1111/j.1574-6968.2000.tb09222.x

Mosley, L. M. (2015). Drought impacts on the water quality of freshwater systems; review and integration. *Earth Sci. Rev.* 140, 203–214. doi: 10.1016/j.earscirev.2014.11.010

Moyer, J., Dakin, R., Hewman, R., and Groves, D. (2009). The case for cyber security in the water sector. *J Am Water Works Assoc.* 101, 30–32. doi: 10.1002/j.1551-8833.2009.tb10007.x

National Science and Technology Council Subcommittee on Ocean Science and Technology (2016). *Harmful Algal Blooms and Hypoxia Comprehensive Research Plan and Action Strategy: An Interagency Report*. Washington, DC: Executive Office of the President Available online at: https://www.whitehouse.gov/wp-content/uploads/2017/12/Harmful-Algal-Blooms-Report-FINAL-August.2017.pdf (accessed May 23, 2019).

Oehrle, S. A., Southwell, B., and Westrick, J. (2010). Detection of various freshwater cyanobacterial toxins using ultra-performance liquid chromatography tandem mass spectrometry. *Toxicon* 55, 965–972. doi: 10.1016/j.toxicon.2009.10.001

O'neil, J. M., Davis, T. W., Burford, M. A., and Gobler, C. J. (2012). The rise of harmful cyanobacteria blooms: the potential roles of eutrophication and climate change. *Harmful Algae* 14, 313–334. doi: 10.1016/j.hal.2011.10.027

Onodera, H., Satake, M., Oshima, Y., Yasumoto, T., and Carmichael, W. W. (1997). New saxitoxin analogues from the freshwater filamentous cyanobacterium Lyngbya wollei. *Nat. Toxins* 5, 146–151. doi: 10.1002/19970504NT4

Otten, T. G., and Paerl, H. W. (2015). Health effects of toxic cyanobacteria in US drinking and recreational waters: our current understanding and proposed direction. *Curr. Environ. Health Rep.* 2, 75–84. doi: 10.1007/s40572-014-0041-9

Paerl, H. W., Hall, N. S., and Calandrino, E. S. (2011). Controlling harmful cyanobacterial blooms in a world experiencing anthropogenic and climatic-induced change. *Sci. Total Environ.* 409, 1739–1745. doi: 10.1016/j.scitotenv.2011.02.001

Panguluri, S., Nelson, T. D., and Wyman, R. P. (2017). "Creating a cyber security culture for your water/waste water utility," in *Cyber-Physical Security*, eds. S. Hakim, E. A. Blackstone, and R. M. Clark, AG (Cham: Springer International Publishing), 133–159. doi: 10.1007/978-3-319-32824-9_7

Powers, C., Hanlon, R., and Schmale, D. G. (2018a). Tracking of a fluorescent dye in a freshwater lake with an unmanned surface vehicle and an unmanned aircraft system. *Remote Sens.* 10:81. doi: 10.3390/rs10010081

Powers, C. W., Hanlon, R., Grothe, H., and Prussin, A. J. (2018b). Coordinated sampling of microorganisms over freshwater and saltwater environments using an unmanned surface vehicle (USV) and a small unmanned aircraft system (sUAS). *Front. Microbiol.* 9:1668. doi: 10.3389/fmicb.2018.01668

Rabalais, N. N., Turner, R. E., Diaz, R. J., and Justić, D. (2009). Global change and eutrophication of coastal waters. *ICES J. Mar. Sci.* 66, 1528–1537. doi: 10.1093/icesjms/fsp047

Sanchez, J. A., Otero, P., Alfonso, A., Ramos, V., Vasconcelos, V., Aráoz, R., et al. (2014). Detection of anatoxin-a and three analogs in *Anabaena* spp. cultures: new fluorescence polarization assay and toxin profile by LC-MS/MS. *Toxins* 6, 402–415. doi: 10.3390/toxins6020402

Sanjab, A., and Saad, W. (2016). Data injection attacks on smart grids with multiple adversaries: a game-theoretic perspective. *IEEE Trans. Smart Grid* 7, 2038–2049. doi: 10.1109/TSG.2016.2550218

Scavia, D., Allan, J. D., Arend, K. K., Bartell, S., Beletsky, D., Bosch, N. S., et al. (2014). Assessing and addressing the re-eutrophication of Lake Erie: central basin hypoxia. *J. Gt. Lakes Res.* 40, 226–246. doi: 10.1016/j.jglr.2014.02.004

Smith, D. R., King, K. W., Johnson, L., Francesconi, W., Richards, P., Baker, D., et al. (2015). Surface runoff and tile drainage transport of phosphorus in the midwestern United States. *J. Environ. Qual.* 44, 495–502. doi: 10.2134/jeq2014.04.0176

Spatharis, S., Tsirtsis, G., Danielidis, D. B., Do Chi, T., and Mouillot, D. (2007). Effects of pulsed nutrient inputs on phytoplankton assemblage structure and blooms in an enclosed coastal area. *Estuar. Coast. Shelf Sci.* 73, 807–815. doi: 10.1016/j.ecss.2007.03.016

Steffen, M. M., Davis, T. W., McKay, R. M. L., Bullerjahn, G. S., Krausfeldt, L. E., Stough, J. M., et al. (2017). Ecophysiological examination of the lake erie microcystis bloom in 2014: linkages between biology and the water supply shutdown of Toledo, OH. *Environ. Sci. Technol.* 51, 6745–6755. doi: 10.1021/acs.est.7b00856

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., et al. (2013). "Climate change 2013: the physical science basis," in *Intergov. Panel Clim. Change Work. Group Contrib. IPCC Fifth Assess. Rep.* (AR5Cambridge Univ Press), 25.

Thees, A., Atari, E., Birbeck, J., Westrick, J. A., and Huntley, J. F. (2018). Isolation and characterization of lake erie bacteria that degrade the cyanobacterial microcystin toxin MC-LR. *J. Gt. Lakes Res.* 45, 138–149. doi: 10.1016/j.jglr.2018.10.013

Triantis, T. M., Kaloudis, T., Zervou, S.-K., and Hiskia, A. (2016). "Determination of microcystins and nodularin in filtered and drinking water by LC-MS/MS," *Handbook of Cyanobacterial Monitoring and Cyanotoxin Analysis*, eds. J. Meriluoto, L. Spoof, and G. A. Codd (Chichester, UK: John Wiley and Sons), 372–378. doi: 10.1002/9781119068761.ch42

Turner, A. D., Waack, J., Lewis, A., Edwards, C., and Lawton, L. (2018). Development and single-laboratory validation of a UHPLC-MS/MS method for quantitation of microcystins and nodularin in natural water, cyanobacteria, shellfish and algal supplement tablet powders. *J. Chromatogr. B* 1074, 111–123. doi: 10.1016/j.jchromb.2017.12.032

Van Esbroeck, C. J., Macrae, M. L., Brunke, R. I., and McKague, K. (2016). Annual and seasonal phosphorus export in surface runoff and tile drainage from agricultural fields with cold temperate climates. *J. Gt. Lakes Res.* 42, 1271–1280. doi: 10.1016/j.jglr.2015.12.014

Vilmin, L., Mogollón, J. M., Beusen, A. H., and Bouwman, A. F. (2018). Forms and subannual variability of nitrogen and phosphorus loading to global river networks over the 20th century. *Glob. Planet. Change* 163, 67–85. doi: 10.1016/j.gloplacha.2018.02.007

Walls, J. T., Wyatt, K. H., Doll, J. C., Rubenstein, E. M., and Rober, A. R. (2018). Hot and toxic: temperature regulates microcystin release from cyanobacteria. *Sci. Total Environ.* 610, 786–795. doi: 10.1016/j.scitotenv.2017.08.149

Westrick, J. A., and Szlag, D. (2018). A cyanotoxin primer for drinking water professionals. *J. Am. Water Works Assoc.* 110, E1–E16. doi: 10.1002/awwa.1088

Wood, S. A., and Dietrich, D. R. (2011). Quantitative assessment of aerosolized cyanobacterial toxins at two New Zealand lakes. *J. Environ. Monit.* 13, 1617–1624. doi: 10.1039/c1em10102a

Wood, S. A., Puddick, J., Fleming, R., and Heussner, A. H. (2017). Detection of anatoxin-producing Phormidium in a New Zealand farm pond and an associated dog death. *N. Z. J. Bot.* 55, 36–46. doi: 10.1080/0028825X.2016.1231122

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers
in Bioengineering and Biotechnology

PERSPECTIVE
published: 26 June 2019
doi: 10.3389/fbioe.2019.00112

Building Capacity for Cyberbiosecurity Training

Lauren C. Richardson*, Stephen M. Lewis and Ryan N. Burnette

Merrick and Company, Arlington, TX, United States

Cyberbiosecurity lies at the intersection of cybersecurity and biosecurity and addresses the protection of valuable biological material and associated information. As an emerging concept, cyberbiosecurity requires the integration of training strategies targeted to both current and future professionals; as well as an increased awareness in the wider stakeholder community. As the discrete discipline of cyberbiosecurity continues to develop, initial training efforts are likely to include workshops and specialized training that bridge the disciplines of information technology (IT) and life sciences. Potential threats, risks, and vulnerabilities will be defined, cooperative relationships formed, and collaborative solutions developed. As the scope of the training framework for assessing potential threats is adapted to various audiences, in-service trainings will ensure awareness and understanding of threats relevant to specific industries. This framework may also be incorporated into existing curricula across IT and science fields. The scope of potential threats is vast, and eventual specialization will likely fall within the realm of IT professionals, who carry the capability for action. In this paper, we identify stakeholders in the development of cybersecurity training; discuss current training methods, educational requirements, and credentialing for professionals in cybersecurity, biosecurity, and life sciences; suggest mechanisms for integration of cyberbiosecurity training into existing training approaches; and discuss potential for future development of specialized professionals.

Keywords: cyberbiosecurity, biosecurity, cybersecurity, training, risk, threat, biosafety, capacity building (including competencies)

OPEN ACCESS

Edited by:
Diane DiEuliis,
National Defense University,
United States

Reviewed by:
Bruno Jorge Antunes Colaço,
University of Trás-os-Montes and Alto Douro, Portugal
Johannes Rath,
University of Vienna, Austria

****Correspondence:***
Lauren C. Richardson
lauren.richardson@merrick.com

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *11 January 2019*
Accepted: *02 May 2019*
Published: *26 June 2019*

Citation:
Richardson LC, Lewis SM and Burnette RN (2019) Building Capacity for Cyberbiosecurity Training. Front. Bioeng. Biotechnol. 7:112. doi: 10.3389/fbioe.2019.00112

INTRODUCTION

Cyberbiosecurity is a new, multidisciplinary concept with potentially significant impacts on the bioeconomy. Cyberbiosecurity addresses the potential for actual malicious destruction, misuse, or exploitation of valuable information, processes, and material at the interface of the life sciences and digital worlds, requiring an understanding of both (Richardson et al., 2019). Though the scope and definition of potential components continues to be refined and expanded, a common language and framework for the training and growth of a cadre of professionals is needed. Here, we propose a potential pathway for the development of cyberbiosecurity training.

IDENTIFYING CYBERBIOSECURITY STAKEHOLDERS

This intersection of cybersecurity and biosecurity has the potential to affect organizations in multiple different fields, from agriculture and manufacturing to healthcare. Though many stakeholders possess a potential interest in the outcomes of cyberbiosecurity, a relatively small subset of individuals are well-suited to its execution. Due to the multidisciplinary nature of this field, those who conduct assessment, protection, and mitigation of cyberbiosecurity vulnerabilities should be well-versed in both the life sciences and information technology.

Today, biosecurity typically falls under the purview of institutional security and biosafety professionals, working together with external biosecurity assessors. It is generally understood that biosecurity requires an understanding of applicable assets (i.e., valuable biological materials and associated data) as well as an understanding of the threat landscape (e.g., negligent scientists without malicious intent, actors with targeted intent of theft or destruction for specific gain, and anarchic disruptors intent on disturbance of the system or organization). Cybersecurity is maintained at an institutional level by information technology professionals with myriad foundational knowledge bases in network security, systems engineering, on-site training in specific protection of systems within an organization.

Developing professionals in this nascent field requires training that draws—at least initially—from disparate disciplines within the life sciences and information technology. An individual with a thorough understanding of information technology and cybersecurity, plus a background in biological sciences, can be taught the basic tenets of risk, threat, and vulnerability assessment to achieve a comprehensive cyberbiosecurity base of knowledge.

STATUS OF CURRICULA IN THE CYBERSECURITY AND BIORISK MANAGEMENT FIELDS

Training in the field of information technology is varied but well-established. While it is certainly possible to become an expert in one of the myriad IT fields via traditional education (i.e., university or vocational training programs), it is not required: many pathways and opportunities exist toward becoming a trained expert in one of the disciplines within the broader field of IT. Depending on the IT discipline in which one wishes to specialize, there are a number of training programs designed to teach individuals with little to no experience. In a traditional academic environment, individuals may choose to major in a computer engineering program, specializing in one of a range of disciplines, from network engineering to software development. Unlike the life sciences, however, IT expertise can also be gained through a less formal path: an individual may choose to attend a training program hosted by a non-academic organization (e.g., Computing Technology Industry Association). Individuals are also able to specialize in cybersecurity through academic and industry training programs. Further, many experts in the field of information technology obtain their primary education and experience through informal, hands-on training in one of a few domains, such as networking, cybersecurity, development, or systems engineering. With respect to cybersecurity training, many professionals also get their experience on the job; that being said, the International Information System Security Certification Consortium (ISC^2) offers both training and a credentialing system aimed at standardizing topics of expertise, including security and risk management, asset security, security operations, security assessment and testing. These subjects are deeply congruent with the training and on-the-job experience offered within the fields of biorisk management and security.

In the life sciences, biosecurity training has historically been comprised of a varied and blended approach of teaching methodologies, including traditional classroom-based, non-traditional classroom-based (e.g., active learning, hands-on workshops), web-based/ online/ on-demand modules, train-the-trainer, on-the-job training, and others. Deciding which training methods to employ rests largely on considerations of (1) expected mastery of content and (2) proficiency of employing the information beyond the training. Further, the type of training content plays a significant role in dictating appropriate training approaches, frequency, and duration.

As science-based disciplines with a backbone in the biological sciences, biosafety and biosecurity benefit from serving a niche community of professionals and students. Like cybersecurity, the foundation of both resides in the broader discipline of risk management; biosafety is in fact a scientific-oriented field of risk assessment, mitigation, and management. Biosecurity, meanwhile, finds its roots in the field of threat assessment and management. Combined, the two disciplines converge at overall biorisk management (Burnette, 2013; Salerno and Gaudioso, 2015). Training in these areas has been largely developed by trade practitioners and official and unofficial repositories of training content; programs are maintained by professional and international organizations (e.g., ABSA International, International Federation of Biosafety Organizations). Effective organization of biosecurity training—as well as the training approaches themselves—remains in development (Minehata et al., 2013; Nixdorff, 2013).

CONVERGENCE OF DISPARATE PROFESSIONAL FIELDS AND CURRICULA

A cyberbiosecurity professional is a practitioner with requisite foundational understanding of biological science principles and practice, fluency in IT lexicon and management, and concept mastery of risk and threat assessment. While this is an appropriate foundation for the cyberbiosecurity professional, additional understanding of a relevant field of practice (e.g., healthcare, pharmaceuticals manufacture) will be required for comprehensive understanding of field-specific vulnerabilities, as well as the ability to develop and promote mitigation strategies to address risks and threats to applicable assets. An ideal candidate may be an individual with a university degree in biology, or they may possess a combination of technical and professional training in IT, post-graduate training in biorisk management, and on-the-job training in cybersecurity. These unique and historically disconnected disciplines are rapidly converging in many sectors. Despite the recent emergence of cyberbiosecurity as a discrete discipline, many professionals possess overlapping skill sets.

END-STATE OF SUCCESSFUL CONVERGENCE

The professional development of specialists in this field will likely be similar to that of similar risk- and threat-based professions in the sciences, such as biosafety or industrial hygiene. Professionals in these fields typically receive training that includes university education in a basic science, post-secondary training in the field, and on-the-job training specific to their organization and position. Though large organizations may be well-served to employ specialized cyberbiosecurity professionals, it is likely most entities will have neither the resources nor the need for full-time employment, and the pool of available personnel will remain small.

Like similar fields, the specialist is not the only player with significant impact: biosafety, biosecurity, and cybersecurity professionals not specialized in cyberbiosecurity will potentially play a much larger role than the specialist, as they will be working on the front lines to recognize and address issues that threaten the science, data, and automation interface with the workforce and the public. Much like the executor of mitigations and corrective actions following a biosecurity audit, these professionals will likely bear the responsibility for following through on necessary measures to ensure sufficient cybersecurity within an organization or facility.

To achieve this end-state, scientific, IT, and security professionals must understand the requirements for and consequences of compliance with protocols and systems directed to address cyberbiosecurity. Just as scientists receive training in and comply with practices and policies to ensure biosafety, biosecurity, and cybersecurity as appropriate to their positions, they should also receive some degree of training in cyberbiosecurity.

STANDARDIZATION OF PROFESSIONAL TRAINING

Standardization of professional training—often resulting in credentialing—is well-established in the fields of biosafety and cybersecurity. ABSA International (formerly the American Biological Safety Association)—in conjunction with the American Society for Microbiology— has developed and maintained dual credentialing programs for biosafety professionals: the Registered Biosafety Professional (RBP) and the Certified Biological Safety Professional (CBSP) programs are experience and exam-based, respectively. Similarly, the International Federation of Biosafety Associations (IFBA) offers credentials in both biosafety and biosecurity. Both of these organizations offer a variety of training programs and curricula that provide foundations toward these credentials. Recently, ABSA International has undertaken an exploratory stance on the development of a biosecurity credentialing program somewhat analogous to the RBP and CBSP. However, with the exception of the IFBA certificate in biosecurity, no standardized biosecurity curricula or credential has been developed and implemented. This is in part due to the fact that biosafety has been a recognized scientific discipline for several decades, during which time biosecurity, as a discrete field of practice, remains inadequately defined.

DEVELOPMENT PROCESS

Need for Awareness and Definition

Today, there is limited awareness of concepts associated with cyberbiosecurity and their potential impacts on the bioeconomy. Significant change is required to move to an end state in which comprehensive management of cyberbiosecurity threats is integrated into existing organizations and systems. In the initial stages, this includes raising awareness regarding risks, threats, and vulnerabilities associated with cyberbiosecurity across many disparate sectors. Professionals within the risk and threat assessment communities are valuable allies and assets in identifying concepts, strategizing for integration, and sculpting the practice of cyberbiosecurity.

Important first steps have been taken in assessment of the potential impacts on the bioeconomy (Murch et al., 2018; Peccoud et al., 2018), definition of the threat landscape (Richardson et al., 2019), and assembly of professionals to begin to describe the field (Murch, 2017). Additional efforts are still needed: some will be described throughout this series, but more discourse is required to define the needs, impacts, and limitations of the field. Though cyberbiosecurity is not a fully-established field, integration of many of the concepts described can be easily integrated into existing training at the universities, technical trainings, and professional continuing education across related fields.

Integration Into Existing Training

It stands to reason that the general field of information technology and cybersecurity is substantially larger that the field of biorisk management. This is especially true with regard to the number of extant professionals, curricula, and credentialing programs; as well as their overall applicability and integration into innumerable industries. In short, IT touches almost every aspect of daily existence. The same is not necessarily true for biorisk management, which remains a highly-specialized field of practice in discrete environments. From this, it can be inferred that incorporating elements of cybersecurity training into life sciences and biorisk management training would be a logical first step toward integrating seemingly disparate curricula. This argument is bolstered in the U.S. by the fact that the U.S. Federal Select Agent Program has stringent requirements surrounding appropriate access of information regarding biological select agents and toxins. This requires institutions with biological select agents and toxins to conform with a certain threshold of cybersecurity [1,2,3]. An achievable goal in support of awareness and definition—as well as in support of building a developing repository of curricula—is to provide existing, relevant cybersecurity training to professionals in the biorisk management field.

Cybersecurity training is widely available via academic and industry-led curricula. There are also myriad massive online open courses (MOOC) available through credible services, some with linkages to universities. Additionally, many universities are offering formal education, including master's programs via remote, online programs; for which accessibility and affordability are key components. As the demand for cybersecurity experts at the intersection of IT and life sciences continues to grow, it is not difficult to imagine that cyberbiosecurity courses will become popular offerings at online and traditional universities.

CURRICULUM DEVELOPMENT

A significant challenge in the burgeoning field of cyberbiosecurity is the development of a curriculum relevant from both discipline and market perspectives. It is reasonable to assume that curricula will be driven by both technical needs (e.g., discrete and relevant content representative of the needs of practitioners) and by the market pool of would-be professionals and students supporting the industry. Given the breadth of existing curriculum in the fields of biosafety, biosecurity, and cybersecurity, it stands to reason that a comprehensive requirements identification process can be conducted to cross-reference the three disparate disciplines at technical and content levels. Further, this requirements development process is likely to reveal substantial information about the market itself. It is anticipated that many independent requirements already in existence (such as biosafety and cybersecurity curricula), will reveal common, logically-linked themes. However, new requirements not currently captured in any singular discipline are likely to be identified; new content will have to be developed to constitute a body of knowledge representative of the field as it is developing today, with a focus on future development.

Like many developing fields, the establishment of instructors, trainers, and teaching professionals capable of maintaining a curriculum focused on industry needs is likely to be one of the more challenging aspects of training in the cyberbiosecurity field; the general lack of professionals who are equally expert in the biological sciences and cybersecurity practices speaks to this challenge. This is also demonstrated by the fact that biosecurity has yet to be adequately codified in the fields it touches (such as laboratories, agriculture, and personalized medicine, among others). The result is a general lack of professionals who can justify their status as a "biosecurity professional." Often, credentials help their respective fields maintain their relevance. For example, we see Registered Biological Safety Professional (RBP) and/or Certified Biological Safety Professional (CBSP) listed as a requirement within job descriptions for biosafety personnel. Accordingly, it is often an expectation that qualified teaching staff have the same credentials. While analogous professional biosecurity credentials will be considered, it is premature to assume this credential will offer any specialization toward cybersecurity.

CREDENTIALING

Credentialing frameworks may need to be designed and implemented in order to identify and educate interested practitioners working within the emerging field of cyberbiosecurity. There is currently a high barrier to learning concepts in each of the cybersecurity and biosecurity disciplines as independent entities; interested parties willing to take the steps toward an applied career in cyberbiosecurity will need to understand the unique challenges that exist within both disciplines. The implementation of credentialing systems may be beneficial toward standardization of the knowledge base required to be an expert in the field. There are also sub-fields of each discipline ostensibly more relevant to the emergence of cyberbiosecurity (e.g., bioinformatics, network security, sequence origin identification, cloud laboratories, machine learning). These sub-fields could be used to develop a credentialing framework (distinct from existing frameworks today) via employing each component as separate training module. Alternatively, it may be prudent for leaders in this emerging field to partner with existing and well-established organizations in cybersecurity credentialing (e.g., International Information System Security Certification Consortium) to develop and implement a cyberbiosecurity training program. At the time of this writing, there is no credentialing system established for biosecurity: it remains in development by organizations like the American Biological Safety Association. Discrete training courses and workshops could also be implemented as a starting point to introduce the need for a credential, as well as to receive support from cybersecurity and biosecurity experts.

THE PATH FORWARD

A new discipline is not built in a matter of months: it grows organically from existing, related fields, and is supported by advocates and experts who recognize its significance and distinction. Additional workshops, papers, and open fora will encourage collaboration for further definition of relevant concepts. Introduction of these concepts should be presented at various professional symposia and conferences in order to raise awareness, introduce ideas for integration, and bring together interested individuals. From these interested parties, a working group may consolidate in order to develop educational materials that can be integrated into professional and academic organizations.

A working group with experts from multiple fields will define gaps and areas for integration across sectors and personnel

[1] (2005). *Possession, Use, and Transfer of Select Agents and Toxins*, in *7 § 331.* United States Code of Federal Regulations.

[2] (2005). *Possession, Use, and Transfer of Select Agents and Toxins, in 9 § 121., United States Code of Federal* Regulations.

[3] (2005). *Select Agents and Toxins, in 42 § 73., United States Code of Federal* Regulations.

within organizations—potentially leading to the development of a formalized cyberbiosecurity curriculum. A well-defined curriculum may be easily integrated into an academic or technical training system, as appropriate. At this stage, a credentialing mechanism is likely to emerge; however, it is challenging to predict whether this curriculum and credentialing system will fall within the scope of the life sciences or IT. Foundational elements of a cyberbiosecurity credentialing framework are currently in development by thought leaders spanning both IT and life sciences disciplines. The biosecurity credential, currently under evaluation by ABSA International, accounts for cybersecurity elements as a basis for and component of an industry credentialing program. Specifications for a credentialing framework will be further developed and include core competencies, such as physical security, regulations and compliance, biorisk management, secure network architecture, identity management, disaster recovery, and security operations in life sciences facilities.

Concepts addressed in cyberbiosecurity span myriad disciplines, so an open dialogue between subject matter experts, as well as affected stakeholders, is required. Professionals in cyberbiosecurity will require not only expertise and training in science and technology concepts, but also the ability to effectively execute a new form of technical communication across disciplines and organizations to achieve comprehensive solutions.

AUTHOR CONTRIBUTIONS

LR, SL, and RB contributed conception and wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

Merrick and Company provided indirect funding support for this paper by way of employment of the authors. No additional funding was provided for this paper by Merrick and Company or any other entity.

REFERENCES

Burnette, R. (2013). *Biosecurity: Understanding, Assessing, and Preventing the Threat*. Hoboken, NJ: John Wiley & Sons.

Minehata, M., Sture, J., Shinomiya, N., and Whitby, S. (2013). Implementing biosecurity education: approaches, resources and programmes. *Sci. Eng. Ethics* 19, 1473–1486. doi: 10.1007/s11948-011-9321-z

Murch, R. S. (eds.). (2017). *Securing the Bioeconomy – Cyberbiosecurity Workshop*. (Arlington).

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol*. 6:39. doi: 10.3389/fbioe.2018.00039

Nixdorff, K. (2013). Education for life scientists on the dual-use implications of their research: commentary on implementing biosecurity education: approaches, resources and programmes. *Sci. Eng. Ethics* 19, 1487–1490. doi: 10.1007/s11948-013-9478-8

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2018). Cyberbiosecurity: from naive trust to risk awareness. *Trends Biotechnol*. 36, 4–7. doi: 10.1016/j.tibtech.2017.10.012

Richardson, L. C., Connell, N. D., Lewis, S. M., Pauwels, E., and Murch, R. S. (2019). Cyberbiosecurity: a call for cooperation in a new threat landscape. *Front. Bioeng. Biotechnol*. 7:99. doi: 10.3389/fbioe.2019.00099

Salerno, R. M., and Gaudioso, J. (2015). *Laboratory Biorisk Management: Biosafety and Biosecurity*. Boca Raton, FL: CRC Press.

Conflict of Interest Statement: LR, SL, and RB were employed by Merrick and Company.

frontiers in Bioengineering and Biotechnology

REVIEW
published: 21 August 2019
doi: 10.3389/fbioe.2019.00182

Cyberbiosecurity Implications for the Laboratory of the Future

J. Craig Reed and Nicolas Dunaway*

Inspirion Biosciences, Frederick, MD, United States

Technological innovation has become an integral and inescapable aspect of our daily existence as almost everything of significance in our world now has a cyber (i.e., relating to, or involving computers, computer networks, information technology, and virtual reality) component associated with it. Every facet of our lives is now touched by technology. As such, we're experiencing a digital transformation. Unfortunately, both as individuals and as a society, we're inadequately prepared to embrace the myriad of vulnerabilities presented by cybertechnologies. Unintended cyber vulnerabilities present significant risks to individuals, organizations, governments and economies. Here, we identify current cybersecurity vulnerabilities found in the life science enterprise and discuss the many ways in which these vulnerabilities present risk to laboratory workers in these facilities, the surrounding community and the environment. We also consider the cyberbiosecurity benefits associated with numerous innovations likely to be present in the laboratory of the future. The challenges associated with cyberbiosecurity vulnerabilities are not insurmountable; they simply require thoughtful consideration by equipment designers, software and control systems developers, and by end users. Organizations and the individuals that comprise them must respect, value, and protect their data. End users must train themselves to look at every piece of laboratory equipment and every process from a cyberbiosecurity perspective. With this approach, cyberbiosecurity vulnerabilities can be minimized or eliminated to the benefit of workers, life science organizations, and national security.

Keywords: biosecurity, cybersecurity, cyberbiosecurity, cyberbiosafety, cyber biorisk management, bioeconomy

OPEN ACCESS

Edited by:
Diane DiEuliis,
National Defense University,
United States

Reviewed by:
Paula Alexandra Oliveira,
University of Trás-os-Montes and Alto Douro, Portugal
Gerald Epstein,
National Defense University,
United States

****Correspondence:***
J. Craig Reed
creed@inspirionbio.com

Specialty section:
This article was submitted to Biosafety and Biosecurity, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *08 March 2019*
Accepted: *11 July 2019*
Published: *21 August 2019*

Citation:
Reed JC and Dunaway N (2019) Cyberbiosecurity Implications for the Laboratory of the Future. Front. Bioeng. Biotechnol. 7:182. doi: 10.3389/fbioe.2019.00182

INTRODUCTION

Containment laboratories in the United States fall within various economic sectors that comprise the bioeconomy: healthcare and medicine, pharmaceuticals, biotechnology, informatics and agriculture. In 2015, these sectors accounted for $4 trillion or 25% of the US gross domestic product (The National Academies of Sciences Engineering Medicine, 2015). There are over 200,000 biological safety level-2 (BSL-2), high containment (i.e., BSL-3) and maximum containment (i.e., BSL-4) laboratories (labs) in the United States

(National Association of County City Health Officials, 2016)[1,2,3,4,5,6,7,8,9]. This includes public and private research, biological production, and diagnostic laboratories. These labs are operated by local, state and federal agencies, academic organizations, and for profit and not-for-profit commercial enterprises. A wide variety of public and private sector containment laboratories fall within the US Department of Homeland Security (DHS) classification of Healthcare and Public Health Sector of our national critical infrastructure[10]. This includes Biological Select Agent and Toxin (BSAT) Program labs, state and local public health labs, blood banks, labs associated with medicine and dentistry, and biological production labs that manufacture biological materials for use as vaccines, medical countermeasures and diagnostic reagents (Department of Homeland Security, 2016). DHS describes critical infrastructure as "… the physical and cyber systems and assets that are so vital to the United States that their incapacity or destruction would have a debilitating impact on our physical or economic security or public health or safety. The nation's critical infrastructure provides the essential services that underpin American society[11]." Information about private sector infrastructure vulnerabilities or data breaches is protected from public release by the Protected Critical Infrastructure Information (PCII) Program if that information is voluntarily shared with the government for the purposes of homeland security[12]. While private sector vulnerabilities are ferreted away, government sector vulnerabilities or data breaches are rarely shared with the public. For example, Title 42. US. Code 262a(h) specifically exempts some information held by the Select Agent program from the Freedom of Information Act[13]. Therefore, while agencies of the federal government have developed awareness of vulnerabilities that exist in these labs, the public, and likely the many individuals who work in these labs, is not apprised of the significant safety and security vulnerabilities present in them[14,15]. This also means that civilian safety and security solution providers cannot use the information that is known about their vulnerabilities to develop solutions[16]. The cyberbiosecurity risks in containment laboratories, discussed below, represent an additional challenge and make an already complicated situation more complex. In short, the footprint is large, the vulnerabilities are significant, and the consequences are high.

CURRENT TECHNOLOGY TRENDS AND THEIR IMPACT TO TODAY'S LABORATORIES

Disruptive technology trends propel the future and the pace of technological innovation is accelerating. There's no question we've entered a period of digital transformation across all aspects of our existence. "Digital transformation is the change associated with the application of digital technologies in all aspects of human endeavor[17]." Through this transformation, technology has become a fundamental aspect of our life. Technology now touches everything of significance in our world and everything of significance now has a cyber component. Of importance, our efficiency and productivity are substantially increased when devices and systems are networked and connected to the internet. This efficiency, in turn, accelerates the pace of disruptive innovation.

Despite massive benefit, technology presents significant security vulnerabilities to the life science enterprise. These vulnerabilities must be managed effectively to avoid existential threat to the enterprise, public health, and national security.

Life science labs are in the early stage of transition to the "smart labs" of the future[18,19,20]. Most existing labs already possess attributes common to residential properties known as "smart homes." Smart homes possess networked devices capable of remote monitoring and control such as thermostats, locks, lighting, televisions, and refrigerators. Users can receive auto-notification of service status (i.e., power on/off) as well as physical changes in the environment such as temperature, motion, or sound. This is similar to networked building automation systems (BAS) and energy management software (EMS) commonly found in modern laboratory facilities. These systems provide climate and humidity control and, importantly, control of pressure differentials between work spaces such as administrative corridors and laboratories that operate at varying levels of containment. When networked, building system performance can be controlled remotely and utility consumption and greenhouse gas emissions can be monitored remotely[21,22]. Some smart systems can schedule recurring preventative maintenance tasks, assign those tasks to specific individuals, and automatically order replacement parts and supplies to maintain stock[23].

[1] http://www.gao.gov/assets/660/652308.pdf
[2] https://report.nih.gov/award/index.cfm
[3] https://news.vin.com/VINNews.aspx?articleId=32051
[4] https://www.naics.com/sic-industry-description/?code=8011
[5] https://www.naics.com/sic-industry-description/?code=8062
[6] https://www.naics.com/sic-industry-description/?code=8069
[7] https://www.naics.com/sic-industry-description/?code=8071
[8] https://www.naics.com/sic-industry-description/?code=8092
[9] https://www.naics.com/sic-industry-description/?code=8099
[10] https://www.dhs.gov/cisa/critical-infrastructure-sectors
[11] https://www.dhs.gov/topic/critical-infrastructure-security
[12] https://www.dhs.gov/pcii-program
[13] https://www.govinfo.gov/content/pkg/USCODE-2010-title42/pdf/USCODE-2010-title42-chap6A-subchapII-partF-subpart1-sec262a.pdf
[14] https://www.usatoday.com/story/news/2015/05/28/biolabs-pathogens-location-incidents/26587505/
[15] https://www.usatoday.com/story/news/2015/05/28/labs-fight-for-secrecy/26530719/
[16] https://money.cnn.com/2015/11/30/technology/secret-deals-hacked-companies/index.html
[17] https://www.shellypalmer.com/events/ces-2018/media-tech-trend-report/
[18] https://www.scientific-computing.com/sites/default/files/content/BASL18%20Web.pdf
[19] http://www.digitaljournal.com/tech-and-science/science/four-pillars-of-the-digital-laboratory/article/506737
[20] https://www.rdmag.com/blog/2016/02/digitally-transforming-laboratory-operations
[21] https://www.csemag.com/articles/networked-bas-energy-management-systems/
[22] https://aquicore.com/blog/building-automation-systems-vs-energy-management-software/
[23] https://www.cxalloy.com/home

Our smart environments at home and work involve networked hardware and mobile communication devices. They are, therefore, subject to the same cybersecurity vulnerabilities. It's widely recognized that hardware and communication devices such as computers and cell phones possess cybersecurity vulnerabilities and once networked, these vulnerabilities can be exploited by anyone with an internet connection. Poor data security and hardware protection habits in one's personal life combined with a remarkable undervaluation of our personal data may translate to similar behaviors and habits in the work environment. Unfortunately, the general consumer does not routinely utilize recommended and proven cybersecurity practices with their personal electronic devices and data. Consumers tend to use short and simple passwords that can be easily guessed, are reused on multiple devices or across multiple accounts and are rarely changed. They conduct financial transactions across open and unsecure public networks. And they give their personal data away for nothing or almost nothing through enrollment and use of loyalty cards at gas pumps, grocery stores, and pharmacies. These poor personal data security habits translate into similar behaviors and practices in the work environment, thus presenting significant cyberbiosecurity vulnerabilities to the life science enterprise.

Life science businesses and academic laboratories rarely respect the value of or take strong measures to protect information about their work environment because they don't realize its sensitivity or appreciate the magnitude of the safety and security vulnerabilities revealed by such documents. Documents such as floorplans for laboratories and mechanical spaces as well as mechanical/electrical/plumbing schematics reveal the location and magnitude of pathogen storage, research animal housing, mission critical reagents, and network servers. They also reveal the identification and location of video surveillance and intrusion detection devices, facility mechanical systems, critical infrastructure components, inbound utility service connections, outbound liquid waste streams, directional airflow and pressure differentials across rooms. To the knowledgeable adversary, every point of information can reveal significant vulnerabilities of the organization. These same organizations may not have restrictions on employee access to this information. Notably, few organizations recognize they lose control of this information once it's distributed to contractors, vendors or service providers such as those who service equipment, manage renovations, or perform space decontamination.

While some life science enterprises may observe other cybersecurity best practices, life science organizations can be complacent about the security of their networked equipment, generally do not properly value their data and business information, and do not fully recognize the significant security vulnerabilities this information may reveal about their organization[24,25,26]. The use of personal devices such as personal laptops and cell phones to access work-related systems results in duplication and redirection of work data streams that introduce additional vulnerabilities and increase the complexity of the cybersecurity challenge for several reasons. First, it requires employers to recognize the necessity to incur the cost associated with either banning the use of personal devices (and issuing company owned devices) or implementing the infrastructure to impose security policies on personal devices that access the organization's networks. Effective cybersecurity policy includes cryptographically strong password usage, use of multifactor authentication, and encryption of data at rest and in transit. While some individuals follow such procedures on their personal devices, most do not. This introduces uncontrolled cyberbiosecurity vulnerabilities to life science enterprise data systems and networked laboratory equipment. Second, personal devices can be used over unsecure public networks—such as in coffee shops or hotel rooms—to access lab systems and data. Without the use of a virtual private network (VPN) or encrypted data, unsecure networks permit other parties to see and intercept transmitted data—a clear vulnerability to any organization. Third, when personal devices are connected to external networks and carried into the lab, they can be used to remove sensitive work data and communicate it to others without detection. Data exfiltration—or data theft—is a perfect example of the insider threat. Fourth, the use of Wi-Fi in a lab or other facility is often a serious vulnerability, and this is exacerbated when allowing personal devices. If a personal device is connected to an organization's internal network and is allowed to broadcast as a Wi-Fi access point, a new point of entry is created for a bad actor. Finally, any mobile device can be lost or stolen. With inadequate security protections, a lost or stolen device can expose the organization's systems and data to intrusion, corruption, and theft. Individuals, businesses, and government agencies are finding that the efficiency and productivity benefits of networking mobile devices, laboratory equipment and facility systems are offset by the crippling security vulnerabilities presented by them. Depending on the size of the organization, remediation of these vulnerabilities can range from a moderately challenging task requiring a single or small number of professionals to a large-scale endeavor requiring a very large team. Regardless, in all cases it requires the organization to implement a risk-based graded approach to information security governance that enables the organization to secure its information, detect loss, and act quickly.

BIOSECURITY VS. CYBERBIOSECURITY

Laboratory biosecurity has been defined as the set of practices and procedures executed at the personal and institutional level necessary to secure and "prevent the loss, theft, misuse, diversion or intentional release of pathogens and toxins" (Meyerson and Reaser, 2002; World Health Organization, 2004). This definition was expanded beyond harmful biological organisms and proteins by Burnette et al. (2013a,b) to include "… products having intrinsic value, such as novel vaccines, biological therapeutics, information technology platforms, synthetic nanoparticles, or

[24] http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.123.9572&rep=rep1&type=pdf

[25] https://blog.societyinsurance.com/common-data-threats-and-vulnerabilities/

[26] https://sloanreview.mit.edu/article/your-data-is-worth-more-than-you-think/

organisms, and products having high monetary value or related to biological agents."

Cyberbiosecurity has been broadly defined by others as "understanding the vulnerabilities to unwanted surveillance, intrusions, and malicious and harmful activities which can occur within or at the interfaces of comingled life and medical sciences, cyber, cyber-physical, supply chain and infrastructure systems, and developing and instituting measures to prevent, protect against, mitigate, investigate, and attribute such threats as it pertains to security, competitiveness, and resilience" (Murch et al., 2018). In this paper we focus our discussion on those aspects of cyberbiosecurity that include all forms of data stored and transmitted through information technology platforms including data streams emanating from networked laboratory equipment, email, electronic documents and files, databases containing sensitive business information, contracts and financial data, raw research data and its analysis, digital inventories of freezer and working stocks, digital genetic and protein sequence, phenotypic and genotypic information about unique recombinant organisms, security access codes, and other intellectual property.

Cyber exploitation of biosecurity vulnerabilities can occur through exfiltration of data by employees or contractors (insiders) or penetration of the organization's networked systems by outsiders. These considerations must be addressed by IT (Information Technology) staff during the collaborative development of a biosecurity program plan (Reed and Sharpe, 2013). Just as the nation's power grid and local utilities are at risk due to the internet accessibility of many individual pieces of networked equipment, so are building automation systems, facility controls and all other networked equipment or communication systems.

Cyber penetration of networked lab equipment and facility controls provides access to the organization's sensitive scientific and business data as well as intellectual property. Aside from denial of service and malware introduction, cyberbiosecurity intrusions and exfiltration of data can result in a cascade of catastrophic reputational and financial outcomes that can challenge the viability of an organization. These outcomes include the destruction, theft, public dissemination of or malicious alteration of electronic genomic and protein sequences, scientific data, intellectual property, and/or security-sensitive facility documents (such as budget documents, program plans, facility floorplans, emergency procedures, continuity of operations plans, etc.). Access to networked laboratory equipment such as freezers, refrigerators and incubators can result in destruction of valuable reagents and microorganisms in long term storage, in use as working stocks, or in active research or experimental use. Networked bench equipment can be turned off and result in lost data and work time. Changes to light, temperature or humidity in animal rooms can result in stress, morbidity or mortality of valuable and expensive research animals. Although we know of no specific events such as these affecting BSAT labs, it is worth noting that only information associated with the loss, theft, release or exposure to Select Agents would be reported to the Select Agent Program —not the destruction of organisms due to a cyberintrusion. The authors are not aware of any requirement forBSL-2 or non-Select Agent BSL-3 labs to report to any authority events such as those described above.

These events can cause irreparable damage to the reputation of individual researchers, principal investigators, specific laboratories, senior leadership of the organization and that of the entire enterprise, institution or federal agency. This, in turn, can erode confidence in the organization by the public as well as current or prospective students, employees, collaborators, sponsors, investors, shareholders and funding agencies. Exploitation of cyberbiosecurity vulnerabilities can be a direct existential threat to the life science enterprise.

CYBERBIOSAFETY AND CYBERBIORISK MANAGEMENT

Cyberbiosecurity is distinguished from cyberbiosafety, which we propose here as a new term for the cyber vulnerabilities associated with networked data systems, laboratory equipment and facility security and engineering controls that may result in environmental contamination or pose a threat to the health of humans, animals and plants including the health of building occupants, the surrounding community, and/or users and consumers of products created by the life science enterprise. Malicious exploitation of cyberbiosafety vulnerabilities include: alteration of electronic genomic sequences to create, enhance or expand infection, host range, pathogenicity or drug resistance of microorganisms (Adam et al., 2011); adjustment of fan speeds in building ventilation systems to alter pressure differentials between administrative and laboratory workspaces which can lead to potential exposure of any building occupant to infectious microorganisms or their toxic products, contamination of the facility, or airborne release of pathogens to the surrounding external environment; and changes to chemical concentration and/or holding time in liquid effluent decontamination systems which can result in premature discharge of infectious, toxic byproducts or genetically altered microorganisms to the municipal waste stream. As with cyberbiosecurity intrusions, the cascade of catastrophic reputational and financial outcomes from cyberbiosafety intrusions represent an existential threat to the life science enterprise and can present a direct immediate threat to the health and safety of building occupants, the public, and the environment.

Although biosecurity, physical security and biosafety are different disciplines, they are synergistic and "… are intimately connected and must be mutually supportive for maximum effectiveness" (Reed and Sharpe, 2013). For any to be fully effective, each must recognize the importance of the other and each must be integrated with the execution of the other. The World Health Organization coined the term biorisk management (World Health Organization, 2006). Biorisk management (BRM) is a management system approach to the identification, elimination and/or mitigation of biosafety and biosecurity risks. We propose here a new term, cyberbiorisk management, as the management system approach to the identification, elimination and/or control of cyberbiosecurity and cyberbiosafety vulnerabilities in the life science enterprise. Detailed discussion of cyberbiosafety and cyberbiorisk management are the focus of a forthcoming publication[27].

CURRENT LABORATORY CYBERBIOSECURITY VULNERABILITIES AND KNOWN ADVERSARIAL EVENTS

One of the more mundane but very real risks to data in the life science enterprise is presented by a single piece of ubiquitous administrative equipment, the all-in-one printer/copier/scanner. This networked device stores vast amounts of unencrypted data received from networked computers through print demands in addition to data created through manual copier and scanner functions. This data is not only vulnerable to theft and misappropriation through cyber penetration, it's also readily accessible when the device is physically or remotely serviced, anytime the data storage is removed, and when the current device is replaced with new equipment. Few organizations stop to consider the massive vulnerability this single piece of equipment presents to the security of all forms of business sensitive information created, handled, and stored by the organization including banking and tax documents, contract terms and scope information, intellectual property, personal identification information, HIPAA (Health Insurance Portability and Accountability Act of 1996)[28] protected information and unpublished research data.

Peccoud et al. (2017) have identified multiple theoretical cyber vulnerabilities associated with networked biomanufacturing process equipment including, supply chain manipulation, alteration of digital genomic sequences, manufacturing process and workflow controls, and the manipulation of process and/or product data. Cyber penetrations that result in alteration of digital genomic or protein sequences could undermine microbial forensics efforts and compromise the ability of the government to distinguish naturally occurring events from deliberate or accidental events. The ability to assign responsibility to malicious actors would be compromised (Reed et al., 2013).

Alteration of processing time and performance of equipment can pose crippling financial and reputational implications due to the loss or destruction of product. On June 27, 2017, the computer networks of the international pharmaceutical company Merck were subject to a global ransomware attack by the NotPetya virus[29,30]. While this attack was not specifically targeted at Merck's biological production or manufacturing control systems, the attack affected international and domestic operations of the company including biologics production of the pediatric vaccine Garadasil (Human Papillomavirus 9-valent Vaccine, Recombinant)[31]. The attack resulted in at least $135 million dollars in lost sales and $175 million in additional costs during the third quarter of 2017 and forced Merck to borrow $240 million worth of Garadasil from the CDC's Pediatric Vaccine Stockpile[32,33]. The attack impacted revenue to the same extent during the fourth quarter of 2017, resulting in a total direct cost to Merck of almost $1 billion. NotPetya racked up more than $10 billion in damages worldwide and has been recognized as the most costly cyber attack in history[34]. Despite the direct financial impact to Merck, it's notable that Merck's forward looking statement of risk found in the fourth quarter 2017 8-K Securities and Exchange Commission filing did not identify cyberbiosecurity issues as a potential risk to shareholders[35].

It's worth noting that future cyber attacks directed specifically at biological production facilities may not only result in the loss or destruction of product, they could potentially result in the creation of potentially harmful products that make their way to end users.

In 2017 at the USENIX security symposium, a group of researchers from the University of Washington presented ground breaking evidence of their ability to encode malware into DNA via a proof-of-concept research project[36]. When the malware-containing DNA was assembled by a gene sequencer, the machine's sequencing software became corrupted. This compromised the computer that controlled the sequencer. Depending upon the networked nature of that computer and the network security protocols in place, this vulnerability could be just the opening an adversary needs to compromise an organization's systems in ways similar to or worse than those described in the paper[37]. This work represents the first demonstration of malicious code insertion into DNA and should be of significant concern to every end user, every gene sequence software developer and every hardware manufacturer.

It's important to emphasize that this work was a proof-of-concept. In phase one of the research, the scientists did not test their theory against a commercially available DNA sequencer/synthesis platform. Instead, the researchers utilized an open source program in which they disabled the security features to create an optimal environment for attack before they introduced the vulnerability. This permitted the researchers to focus their attention solely on the biochemical challenges associated with DNA-based cyber exploitation. Further, the vulnerability introduced by the group was a "buffer overflow." It is easy to focus on the artificiality of the engineered vulnerability and, perhaps, conclude this somehow invalidates the research. Some might even conclude that this research

[27] Reed, manuscript in preparation.

[28] https://www.govinfo.gov/content/pkg/PLAW-104publ191/html/PLAW-104publ191.htm

[29] https://www.washingtonpost.com/news/the-switch/wp/2017/06/27/pharmaceutical-giant-rocked-by-ransomware-attack/?noredirect=on&utm_term=.4a9d7b51fc4b

[30] https://www.fiercepharma.com/manufacturing/merck-says-its-has-restored-most-its-manufacturing-hit-by-cyber-attack

[31] https://www.apextechservices.com/topics/articles/435235-notpetya-worlds-first-10-billion-malware.htm

[32] https://www.cyberscoop.com/notpetya-ransomware-cost-merck-310-million

[33] https://www.techrepublic.com/article/notpetya-ransomware-outbreak-cost-merck-more-than-300m-per-quarter/

[34] https://www.apextechservices.com/topics/articles/435235-notpetya-worlds-first-10-billion-malware.htm

[35] https://fintel.io/doc/sec/310158/000110465918006007/a18-5152_18k.htm

[36] https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/ney

[37] https://www.usenix.org/system/files/conference/usenixsecurity17/sec17-ney.pdf

should have used a novel vulnerability to be useful and that the scenario created by the researchers was so artificial so as to have no intrinsic value. However, this ignores several key points.

First, the use of a buffer overflow as the vulnerability (however artificially engineered), was an interesting choice because buffer overflows have been documented as early as 1972 and are not only one of the oldest known cyber vulnerabilities[38] but one which often remains unaddressed in modern software releases today[39]. The choice of vulnerability was wise, as it reveals that many software developers over the years have not placed (and still do not place) appropriate priority on security hygiene when engineering their code base. Second, the University of Washington researchers demonstrated this in phase two of their research through their interrogation of poor security hygiene practices in commonly used next-generation sequencing (NGS) and bioinformatics programs. The researchers identified many vulnerabilities, including several buffer overflow vulnerabilities, across different programs.

Simply stated, focusing solely on the vulnerability the researchers chose to exploit may cause some to overlook the critical lessons. Namely, that it is possible (in some cases) to encode malware into DNA, and that many NGS and bioinformatics programs utilize poor security hygiene practices.

Frankly, it's no surprise that life science software developers generally give little to no priority to security hygiene considering that the overall security hygiene in traditional software code is poor. It's imperative this trend be reversed. Manufacturers of NGS, bioinformatics software, and all life science software must consider cyberbiosecurity at the outset of product development, not as an afterthought or with the absence of thought.

Concerns about the cyber vulnerabilities of mobile medical devices captured the public's attention in 2012 when a popular television drama depicted a pacemaker assassination attempt of a fictious political figure[40]. That same year, the US Government Accountability Office identified multiple security vulnerabilities associated with mobile medical devices which are also significant to the life science enterprise (United States Government Accountability Office, 2012), including: unsecure access, unencrypted data transfer, and an inability to update or install security patches or software updates. These vulnerabilities have been exploited by one individual to program an artificial heart to produce a lethal 830 volt shock[41] and to reprogram an insulin pump to release sufficient insulin to kill its wearer without any warning[42]. In 2013, it was revealed that the unsecured wireless capability of Vice President Dick Cheney's defibrillator was disabled due to the possibility it could be remotely inactivated. The wireless function was intended for software updates to the device[43].

DISRUPTIVE TECHNOLOGY AND CYBER VULNERABILITIES IN THE LAB OF THE FUTURE—WELCOME TO THE SMART LAB

The Lab of the Future (LotF) will be known as a "smart lab"—a concept simultaneously exhilarating and daunting. "Exhilarating" because the injection of disruptive technology into this workspace will further accelerate the pace and efficiency of innovative research that will, in turn, improve human health, our quality of life and longevity. "Daunting" because the disruptive technology introduces human health risks and security vulnerabilities which must be anticipated, carefully evaluated, and thoughtfully mitigated. Already existing cyber vulnerabilities coupled with the challenges associated with integrating and securing new and disruptive technology may explain why disruptive technology trends have been slower to enter the laboratory workspace than our personal lives. But a number of consumer electronic trends suggest that, ultimately, the LotF will integrate and fully embrace the very same technologies we find ourselves using today at home, as well as those that are just over the horizon.

Lab of the Future Driven by Virtual Personal Assistants

The portal to the LotF is the voice-driven (or virtual) personal assistant (VPA). According to The Palmer Group, "the world is increasingly mobile and connected[44]." This same organization identified on-demand services as one of technology's current megatrends. "People are not only willing to access goods and service when they need them, they are getting used to living in a world where their demands are instantly met" (2018 Media & Tech Trend Report). These demands are increasingly met through voice activation services which have now become mainstream because consumers prefer voice activation to typing commands. It's worth noting that speech recognition is 3x faster than typing on smart devices (Ruan et al., 2017).

Siri, Cortana, Google Assistant, and Alexa are some of the most well-known VPAs and are created, respectively, by Apple, Microsoft, Google, and Amazon. Siri, Cortana, and Google Assistant are designed primarily for use on mobile phones or other computer platforms. Alexa is a VPA designed to primarily function inside the Amazon Echo series of smart speaker and video devices. Smart speakers are always-on internet connected devices possessing speakers and omni-directional microphones. Their primary input and output are voice (or voice and video for video enabled devices). Through natural language processing (NLP), location data and access to cloud-stored data, these devices provide audio information directly to users and allow users to access, control and monitor internet-connected

[38] http://csrc.nist.gov/publications/history/ande72.pdf

[39] https://securityaffairs.co/wordpress/60507/hacking/skype-buffer-overflow.html

[40] https://www.forbes.com/sites/singularity/2012/12/06/yes-you-can-hack-a-pacemaker-and-other-medical-devices-too/#4bcaa81d6853

[41] https://www.computerworld.com/article/2492453/malware-vulnerabilities/pacemaker-hack-can-deliver-deadly-830-volt-jolt.html

[42] https://www.theregister.co.uk/2011/10/27/fatal_insulin_pump_attack/

[43] https://nakedsecurity.sophos.com/2013/10/22/doctors-disabled-wireless-in-dick-cheneys-pacemaker-to-thwart-hacking/

[44] https://www.shellypalmer.com/events/ces-2018/media-tech-trend-report/

products such as thermostats, lighting, security systems, and household appliances. Natural language processing is one of five subdomains of artificial intelligence (AI) and is the technology that enables a computer to both understand and respond in any human language. Natural language processing is what enables a VPA to receive spoken directions and respond with a human voice[45].

Amazon's smart speakers sold over 22 million units in 2017 and are projected to have a US household adoption rate of 55% by 2022[46,47]. At this rate, Amazon's smart speakers will become the fastest-adopted consumer electronics device in history[48]. Owners of smart speakers can't do without them—50% use them daily[49] and more than 30% of owners have more than one device[50]. Park Associates observes, "… voice interface creates a natural gateway to smart home products with consumers desiring to build their ecosystem around voice, thus leading to greater smart home adoptions. [Our] research supports this strong correlation between smart home ownership and adoption of smart speakers with personal assistants[49]." The ease of voice-based services, combined with the rapid adoption of smart speakers in the consumer market portends an abundance of these devices in the workplace and throughout the life science enterprise.

With such massive numbers for only one popular VPA, it is easy to see how this trend will transfer to the lab. Just as home-based smart speakers are placed throughout the home and used to play music, order pizza, or call a parent; it is really just a matter of time before lab-based smart speakers will be used for similar functions in administrative spaces and laboratories of the scientific enterprise. Smart speakers will be unobtrusively mounted throughout the life science complex in the walls and ceilings of rooms, corridors and laboratories. When flat panel monitors are mounted and networked in the laboratory, conference rooms or huddle rooms, users will request smart speakers to present standard operating procedures, training videos, written documents, and electronic laboratory notebooks (ELNs) on demand. To reduce or eliminate disruption of others in the workplace, smart speakers will be paired with Bluetooth enabled earbuds to enable discrete communication with and receipt of audio content from the networked system of speakers. Individuals will be able to use smart speakers to notify leadership of security and safety emergencies.

Consumer equipment manufacturers wishing to support VPA interactivity currently use the appropriate software development kit (e.g., Alexa Skills Kit, Apple Development for HomeKit, Actions on Google, etc.) to enable their equipment to electronically interface with the NLP capabilities of the VPA. This software is embedded in the software of household smart devices such as lightbulbs, locks, thermostats and refrigerators in "smart homes[51,52,53]." In the future, software developed by laboratory equipment manufacturers will permit scientists and technicians to use voice commands through smart speakers to control and monitor networked laboratory equipment (i.e., centrifuges, incubators and biosafety cabinets) and data generating equipment (i.e., sequencers and plate readers). The increased use of smart speakers and the expansion of skills will decrease the need for printed documents in the laboratory and accelerate electronic laboratory notebook (ELN) adoption as scientists use smart speakers to dictate select information into e-notebooks, direct the import of data streams from networked bench equipment, and interact with laboratory information management systems. While physical lab notebooks are portable and can be misplaced, lost, damaged or destroyed, ELNs are more secure because they can be encrypted, password protected and stored in the cloud. ELNs that meet regulatory requirements and include the appropriate audit trail and e-signature features may also enhance laboratory quality management, including compliance with Good Laboratory Practices and Good Manufacturing Practices (Kwok, 2018). Software developed by BAS designers will permit facility engineers to remotely monitor and adjust the performance of facility systems. Voice-activated systems and equipment will improve worker productivity and efficiency just as they do in our personal lives. They'll also potentially decrease the likelihood of infection and work surface contamination due to decreased touch in the work space.

Voice Biometric Authentication as Part of Multimodal Biometric and Multifactor Authentication Improves Security

Current innovations in voice biometric authentication systems (VBAS) combined with smart speaker ease of use will propel smart speaker adoption in the life science workspace. Voice biometric authentication systems such as those currently used by Homeland Security at border crossings[54] and in the financial, insurance and information technology industries permit the unique identification of individuals based upon their voiceprint[55,56]. Voiceprints are created from over 100 unique physical and behavioral characteristics that contribute to tone, frequency and cadence of an individual's voice[57,58]. As the ability

[45] https://towardsdatascience.com/how-amazon-alexa-works-your-guide-to-natural-language-processing-ai-7506004709d3

[46] https://www.forbes.com/sites/gilpress/2017/10/29/22-million-amazon-echo-smart-speakers-to-be-sold-in-2017-driving-us-smart-home-adoption/#2bcd180b481a

[47] https://techcrunch.com/2017/11/08/voice-enabled-smart-speakers-to-reach-55-of-u-s-households-by-2022-says-report/?utm_medium=TCnewsletter

[48] https://adage.com/article/opinion/amazon-alexa-spying/313672/

[49] http://www.parksassociates.com/bento/shop/whitepapers/files/Parks%20Assoc%20-%20Impact%20of%20Voice%20Whitepaper%202017.pdf

[50] https://www.forbes.com/sites/gilpress/2017/10/29/22-million-amazon-echo-smart-speakers-to-be-sold-in-2017-driving-us-smart-home-adoption/#2bcd180b481a

[51] https://developer.amazon.com/alexa-skills-kit

[52] https://developer.apple.com/homekit/

[53] https://developers.google.com/actions/

[54] https://www.globalsecurity.org/security/systems/biometrics-voice.htm

[55] https://biztechmagazine.com/article/2018/11/voiceprint-security-game-changer-banks-and-credit-unions-all-sizes

[56] https://identity.utexas.edu/assets/uploads/publications/Current-Biometric-Adoption-and-Trends.pdf

[57] https://www.theguardian.com/money/2018/sep/22/voice-recognition-is-it-really-as-secure-as-it-sounds

[58] http://www.nuance-media.eu/sites/default/files/pdf/Voice%20biometrics%20FAQ%20Press%20614.pdf

of VBAS to rapidly and reliably distinguish individuals increases, it will drive this technology to become an integral component of cyberbiosecurity. When the security enhancement associated with VBAS becomes deployed in the marketplace, many pieces of equipment that are or can be networked in the life science enterprise will benefit from this technology.

Currently, smart speakers and VPAs on the market are user agnostic. While they recognize voice commands, they're unable to distinguish the individuals who issue those commands. Once this limitation is overcome and individual users can be distinguished, integration of VBAS into smart speakers will permit the VPA to distinguish unauthorized users from authorized users and to parse commands based on a user's security authorizations. Once VBAS is integrated, smart speakers will provide life science organizations much greater control of physical security and cybersecurity.

We note that VBAS is not a panacea. It is but one aspect of authentication and by itself is insufficient to provide proper security. To understand why, we must first review the three common categories of authentication. First, is "something you remember or know," such as your traditional password (Kumar and Farik, 2016). Second, is "something you possess" (Kumar and Farik, 2016). A centuries-old example is a key used to open a lock. Modern examples might include smart cards, software tokens or other hardware devices. Third, is "something you are" (Kumar and Farik, 2016). This includes all biometrics including voiceprint but also includes "…fingerprint, face, iris, retina, gait, palm, and many more…" (Kumar and Farik, 2016).

A fundamental principle in security is the use of multifactor authentication (MFA). Multifactor authentication is the requirement to use two or more forms of authentication to verify the identity of an individual. For example, if an individual wants to log on to a computer or enter a restricted space through a locked door, the individual is required to present something they "have" (such as a smart card), and something they "are" (such as a voiceprint). Alternatively, they could present something they "know" (a password) and something they "are" (a fingerprint). The system does not have to be limited to two factors, it can easily require three or more. In fact, biometric authentication is precisely what many security experts recommend. Because of the limitations of some forms of biometric authentication (such as VBAS) and the ease with which multiple biometric factors ("multimodal biometric authentication"; Kumar and Farik, 2016) can be paired, and the greatly increased security gained by such paring, it is vital VBAS not be discarded by those who only see the risks. Instead, it should be embraced for its ability to enable the life sciences to operate in a radically more efficient environment while remaining safe and secure.

With VBAS using multimodal biometric, and multi-factor authentication, VPAs will not only be able to restrict an individual's physical access to parts of the building, the VPA will also restrict access to networked equipment. It will also support implementation of the organization's IT Security Plan by maintaining access control over business sensitive documents stored on the organization's servers. VPAs with multimodal biometric and multifactor authentication will serve to greatly enhance the security posture of any organization wise enough to employ them.

Wearables to Monitor Human Performance in High Risk Environments

A relatively undiscussed aspect of containment laboratory operations is physiology, psychology, and human performance monitoring. High containment laboratories (biosafety level-3; BSL-3) and maximum containment laboratories (biosafety level-4; BSL-4) present risks to workers, public health, the environment and national security. Physical medical conditions and mental health issues can impair the ability of individuals to work safely and securely in these environments. Blood sugar imbalances affect dexterity, fine motor skills, vision, balance, clarity of thought, emotional state and executive function[59,60]. A 2017 report from the US Centers for Disease Control and Prevention (CDC) reveals that 30% of the US population is insulin resistant and displays higher than normal blood glucose levels; they are pre-diabetic. An additional 9.4% of the US population is diabetic[61]. Anxiety disorders—the most common mental illness in the United States—affect 18% of US adults[62,63]. Additionally, 18% of individuals between the ages of 45–64 were prescribed antidepressants between 2011 and 2014 (Yan, 2017).

We are aware of only two behavioral health screening processes associated with worker access to high and maximum containment laboratories. To assess and monitor the medical and psychological suitability of individuals to work in high and/or maximum containment environments, the U.S. Department of Defense operates a Biosurety Program which establishes a Biological Personnel Reliability Program [BPRP; Department of the Army (2008)]. The BPRP requires medical screening, evaluation, and certification of individuals who have access to BSAT. A medical evaluation is performed to verify candidates are "… free of unstable medical conditions … drug/substance and alcohol abuse and/or dependence …" Disqualifying factors include alcohol-related incidents, alcohol abuse, drug/substance abuse, and "any significant mental or physical medical condition, medication usage, or medical treatment, which may result in … an altered state of consciousness … impaired judgment or concentration." Individuals are subject to a mental health assessment, as well, and can be disqualified for "… attempted or threatened suicide … extreme moods or mood swings … aggressive/threatening behavior toward other individuals." Maximum containment laboratory workers at the National Institutes of Health, Bethesda, Maryland are also subject to behavioral health screening (Skvorc and Wilson, 2011). Because activities performed in high and maximum containment laboratories potentially pose unique threats to public health and national security, human performance monitoring through the

[59] https://www.webmd.com/diabetes/guide/diabetes-hypoglycemia#1-2

[60] https://www.mayoclinic.org/diseases-conditions/diabetes/symptoms-causes/syc-20371444

[61] https://www.cdc.gov/diabetes/data/statistics/statistics-report.html

[62] http://www.mentalhealthamerica.net/issues/state-mental-health-america

[63] https://adaa.org/about-adaa/press-room/facts-statistics

use of digital health products known as "wearables" may be useful in the future.

The consumer market contains a bevy of wearables including fitness trackers, heartrate monitors and glucose monitors. These devices have become extremely sensitive and permit continuous monitoring of a variety of physiological conditions. For example, the Apple Watch does more than simply monitor heart rate. It can also detect heart arrhythmia with 97% accuracy and hypertension with 82% accuracy[64]. Other devices can monitor blood glucose levels without compromising skin integrity[65,66]. Approximately 125.5 million wearable devices were sold in 2017 and 240 million are projected to be sold in 2021[67]. When these devices are networked it becomes possible to remotely monitor the vital signs, metabolic status and overall physiological state of individuals[68,69]. This could prove helpful, for example, for individuals with diagnosed and undiagnosed medical conditions such as hypoglycemia, diabetes, pre-diabetic syndrome, cardiovascular disease or heart arrhythmia. Permission will certainly be required from the individual to be monitored and the resultant data will be subject to protection under the HIPAA. It is reasonable to anticipate that wearable devices such as these will not only be found in the high and maximum containment lab of the future, but also in other work environments, as well. Some will not agree to the value or to the collection of this information and, instead, will view it as an unacceptable invasion of privacy. On the other hand, wearable technologies to monitor human performance are widely used in elite athletic training. However, logical evaluation of the history of data breaches and medical device cyber vulnerability (some of which we have described) could lead one to have concerns. These concerns further reinforce the need for organizations to provide robust, transparent, cyberbiorisk protections to alleviate the vulnerabilities associated with these new technologies. Discussion about the application of wearable technologies in the containment environment are likely to gain acceptance in the future due to the safety and security implications to the individual, environment and society.

Virtual Reality

Virtual reality (VR) will become a valuable training asset in the lab of the future. Virtual reality replicates or creates an environmental space and is, therefore, perfectly suited for the creation of an exquisitely controlled and focused environment conducive to training. With current technology, trainees can don VR headsets and utilize controllers with basic haptic feedback, to become immersed in a completely safe, risk-free environment, where they will learn by doing. Very soon, more advanced haptic devices such as gloves[70,71,72] will allow for fully immersive and complex activities. With these training devices available, trainees can be objectively scored on their ability to successfully perform activities like donning/doffing personal protective equipment, preparing a biological safety cabinet for work, or disinfecting a biosafety cabinet following work activities. Training will be self-correcting, through the application of game design principles in non-gaming contexts, defined in Robson et al. (2015) as "gamification," where trainees increase proficiency through repetitive rounds of practice combined with advancement and reward systems that incentivize progress (Hao and Chuen-Tsai, 2011). Applications that lend themselves to this type of approach include practical testing during biosafety cabinet field certification, training of animal handlers, and BSAT handling activities. With the integration of haptic devices and various hardware components [i.e., gloves, PAPR (Powered Air-Purifying Respirator), bonnet etc.], trainees will be able to sense temperature, vibration, and texture. This will become an essential aspect of training related to animal handling, use of sharps involving animals and activities both delicate and dangerous involving BSAT.

Virtual reality may eliminate the need for trainers or trainees to travel for training events. Instead, trainers will ship VR devices to trainees pre-loaded with instructional software and training content. Trainees will simulate the performance of training tasks at a location, time, frequency, and tempo of their choice. Like a video game, trainees will be able to repeat the simulated events as many times as necessary to move through successively higher levels of achievement to reach their desired level of proficiency. In the end, trainees will demonstrate greater competency in less time through engagement in a fully immersive learning process they can control.

Artificial Intelligence

According to The Palmer Group, machine learning is one of three megatrends in the field of consumer technology (on demand and autonomy are the other two). "From simple algorithms to complex neural networks, machines are learning to think with us and for us. No matter what you do, there's a thinking machine in your future[73]."

The implications for VPAs and smart speakers go far beyond touch-free work spaces, voice activation of equipment, document display, and notification of safety/security representatives. The full value of VPAs and smart speakers will be unlocked when scores of organizations network their smart speakers to create a higher order system. At this point, the full digital transformation of the LotF will be underway. Individual pools of data from participating organizations will form a data lake of enough size to permit the application of largescale data analysis, machine learning (ML) algorithms, and artificial intelligence (AI).

AI is the broad science of training a machine to emulate human abilities to perform human tasks[74] (Turing, 1950;

[64] https://techcrunch.com/2017/11/13/the-apple-watch-can-accurately-detect-detect-hypertension-and-sleep-apnea-a-new-study-suggests/?utm_medium=TCnewsletter

[65] http://www.gluco-wise.com/

[66] http://nemauramedical.com/sugarbeat/

[67] https://www.digitaltrends.com/mobile/idc-wearables-maket-2017/

[68] https://www.computer.org/csdl/magazine/mu/2018/01/mmu2018010061/13rRUwInvc3

[69] https://www.computer.org/publications/tech-news/research/wearables-smart-phones-sensing-technologies-mental-illness

[70] https://lmts.epfl.ch/lmts-research/blindpad/dextres-2/

[71] https://haptx.com/

[72] https://www.vrgluv.com/

[73] https://www.shellypalmer.com/events/ces-2018/media-tech-trend-report/

[74] http://jmc.stanford.edu/articles/whatisai/whatisai.pdf

Shannon and McCarthy, 1956; McCarthy et al., 2006; Shubhendu and Vijay, 2013). AI encompasses numerous subfields: machine learning (ML), speech, expert systems, computer vision, robotics, planning/scheduling/optimization, and natural language processing (NLP)[75]. Machine learning applies various forms of data analysis to massive volumes of highly granular and diverse pieces of data to identify broad patterns and draw conclusions (Singh et al., 2016). During the process of data analysis, the massive reservoirs of data are inspected, cleaned, transformed, and modeled to identify useful information, draw conclusions and inform decision-making processes.

Increasingly powerful forms of data analysis require increasing volumes of data ("big data") and greater computational resources. These analytic processes are called: descriptive, diagnostic, predictive, prescriptive, and cognitive analytics, with cognitive analytics providing insight and outcomes of the greatest value and use. Descriptive analytics draws upon the mining of comprehensive historical and live data to answer the question, "What happened?" (Banerjee et al., 2013). Diagnostic analytics applies cause and effect analyses to identify correlations within data to enable the isolation of confounding information and identification of root cause, thereby answering the question, "Why did it happen?" (Banerjee et al., 2013). Predictive analytics employs algorithms to identify historical patterns and to generate and assess theoretical models to yield predictive forecasts that answer the question, "What could happen?" (Banerjee et al., 2013). Prescriptive analytics is the application of advanced analytical techniques including simulation, optimization and decision modeling to generate best possible recommendations to answer the question, "What should be done?" (Banerjee et al., 2013). Cognitive analytics, in the realm of AI, prompts action or causes something to be done (Gudivada et al., 2016).

The combined application of data analytics, ML and other AI tools to a large and ever-increasing volume of data enables for example, Amazon and Netflix to not only generate personalized recommendations for consumers based upon their individual browsing and purchase behavior, but also to accurately forecast what products and media content will move fastest with any given demographic and to predict when consumers are likely to demand it[76,77].

Just as every driverless car in a networked fleet learns from the individual mistakes of every other car in that networked fleet, ML and AI may enable individual laboratories to learn from and avoid the errors, incidents, and accidents that have occurred in any other networked laboratory. If one laboratory notifies the safety office of an incident or an accident, AI will have the ability to analyze the information and enable other laboratories to avoid the issues that led to the incident or the accident. Although massive volumes of data are required for data analytics and AI tools to become effective, once sufficient historical and live laboratory data has been amassed, ML tools will assist institutional security and safety committees in the identification and correction of vulnerabilities associated with administrative controls such as standard operating procedures, animal care and welfare procedures, and security processes. With repeated use, the quality of diagnostic and predictive analytics outcomes and the recommendations produced through prescriptive analytics will be refined, their value will increase, and laboratories will be become more reliant upon them. Ultimately, AI generated recommendations and advice will be provided directly to laboratory staff in real time through the smart speaker or other VPA enabled device to prevent unsafe practices that may cause imminent harm to workers. Connection of video feeds and other security sensors to the system is likely to enable security forces and building occupants to be warned of an imminent security threat.

The application of ML and other AI tools to analyze historical laboratory data, data streams from bench equipment, data from wearables, the online behaviors of users on the organization's computers, downloads, and print demands of business documents, email content, and digitally recorded phone conversations will result in revelatory insights about worker behavior, safety practices, and security procedures. The previous and current actions, behaviors, and physiological conditions associated with the insider threat will be apparent for senior leadership as well as security and safety professionals to recognize. Data analysis of news media—both print and online—as well as insurance case studies, and transcripts of court proceedings and judgements for laboratory-related security and safety claims and awards will be useful in learning from past laboratory accidents and intrusions.

For all the potential helpful benefits that ML and AI tools may bring to the laboratory, there will be well-founded concerns associated with the accessibility and security of the raw data from individual laboratories as well as the aggregated raw data. This information could be used to make inferences about the specific activities, operational details and/or safety and security vulnerabilities associated with a given organization and, potentially, with specific subordinate laboratories.

Blockchain Technology

Blockchain has been hailed as "… the most important invention since the Internet itself" and is "… an invention like the steam or combustion engine that has the potential to transform the world …[78]." What makes blockchain powerful is that it works flawlessly and has done so for over a decade as the backbone for cryptocurrencies such as bitcoin[79]. Blockchain can be explained in several ways, some more technical than others. In simple terms, blockchain is a digital audit trail; a shared electronic ledger of all transactions and digital events that is simultaneously secure and verifiable. The "chain" itself is composed of individual "blocks"—or transactions—each of which is attached to the chain in temporal fashion immediately following verification of the current transaction by a majority of

[75] www.cse.scu.edu/~mwang2/ai/AI_subfields.pdf

[76] https://www.predictiveanalyticsworld.com/patimes/how-netflix-uses-big-data-to-drive-success/9693/

[77] https://www.investopedia.com/articles/insights/090716/7-ways-amazon-uses-big-data-stalk-you-amzn.asp

[78] https://scet.berkeley.edu/wp-content/uploads/BlockchainPaper.pdf

[79] https://bitcoin.org/bitcoin.pdf

users. The blockchain contains a certain and verifiable record of every single transaction ever made in the chain. Accountability is, therefore, 100%. Falsification of or tampering with the transaction is not possible in a blockchain[80]. This is due to the distributed consensus model of the blockchain (as compared to say a traditional centralized database model in which trust in the database requires trust in the entity maintaining it). Because all parties involved with every transaction are recorded in the blockchain, it is impossible for anyone to execute an unrecorded transaction. Therefore, blockchain eliminates the untraceable insider (and outsider) threat. This is a significant distinction from even the most stringent internal security protocols that do not use blockchain. Such systems will always have a weak point, such as a system administrator or executive level security officer who could theoretically bypass security protocols to "cover their tracks." With the use of blockchain, this is not possible. This should be the most significant aspect of blockchain from an organization's internal cybersecurity and cyberbiosecurity perspectives.

As blockchain technology has become more popular, it has become marketed as a one-stop solution to cyber security challenges. This is misguided at best and can only serve to slow the adoption of this powerful technology. It is important for users to not only understand what blockchain is and what it can do, but just as importantly, what blockchain is not and what it cannot do (by itself). Blockchain is a means to track transactions with utmost confidence. Blockchain can be used to secure private and confidential information such as intellectual property, security plans, or other sensitive data, however blockchain cannot do this by itself. By using blockchain technology as a layer of security (as described above), other security technologies (encryption of data in transit and at rest, multimodal biometric, multi-factor authentication, and many others) can be employed with the confidence that their use will be tracked (by the blockchain) without worry of alteration or manipulation. In other words, by leveraging blockchain technology, the rest of the security systems in place for an organization will be improved and therefore, whatever those security systems are protecting, will likewise see improved protection. It is worth noting this still requires knowledgeable and dutiful security professionals at the helm to manage, monitor, and respond to the data being tracked by the blockchain.

There are myriad financial and non-financial applications for blockchain technology, some of which are already starting to emerge. For example, DHL reported on the potential for blockchain to protect their global logistics[81], SAP (a German multinational software corporation with over $28 billion of revenue in 2018) is tracking goods from creation to shipment using blockchain to similarly protect their supply chain[82,83] and Kodak is using blockchain to digitally protect intellectual property for photographers[84]. In the life sciences, any transaction (digital or physical) that incorporates blockchain technology will gain the benefits of accountability and validation inherent with this technology. Blockchain can be used for chain of custody for BSAT and other high consequence materials, tracking and authentication of laboratory supply chains, authentication of waste management processes, tracking of sensitive documents (digital and physical) and verification of digital genomic and protein sequences. For the benefit of having secure, verifiable and immutable transactional data, blockchain technology can and should be integrated into cyberbiorisk management software solutions of the future.

RECOMMENDATIONS & CONCLUSIONS

The US government has not issued regulations focused on private-sector computer network security, aside from healthcare and financial data laws enacted in 1996 and 1999, even though 90% of US cyberspace infrastructure is owned and operated by private companies and represents the first line of defense in a cyberwar. Instead, the government encourages voluntary improvements to cybersecurity practices saying simply, "The majority of intrusions can be stopped through relatively basic cybersecurity investments that companies can and must make themselves[85]." To this end, the National Institutes of Standards and Technology has created voluntary computer security guidance to decrease the vulnerability of and increase the resiliency of commercial sector enterprise in the event of a cyber attack (National Institute of Standards Technology, 2018). A 2014 report focused on life science data security conducted by the American Association for the Advancement of Science, the Federal Bureau of Investigation and the United National Interregional Crime and Justice Research Institute came to a similar conclusion, stating: "When evaluating solutions for reducing the vulnerabilities of Big Data in the life sciences, only technical solutions, including access controls and data encryption, exist … Unfortunately, beyond the use of technical solutions and common-sense behavior, institutions and individuals can do very little to address system vulnerabilities" (Berger and Roderick, 2014). While no cybersecurity system is completely impenetrable, this should not preclude individuals or organizations from utilizing proven practices to protect their systems and assets. Doing so makes the task of exploitation more difficult and will likely send an intruder to seek easier targets.

The cyberbiosecurity vulnerabilities of the scientific enterprise identified here and elsewhere can be attributed to several fundamental causes:

- Failure to respect, value and protect the organization's scientific data and business sensitive information;

[80] https://scet.berkeley.edu/wp-content/uploads/BlockchainPaper.pdf

[81] https://newsroom.accenture.com/news/dhl-and-accenture-unlock-the-power-of-blockchain-in-logistics.htm

[82] https://www.nasdaq.com/symbol/sap/financials?query=income-statement

[83] https://www.computerworld.com/article/3298578/sap-pilots-blockchain-based-supply-chain-tracker.html

[84] https://www.digitaltrends.com/photography/kodakone-creates-photo-registry-blockchain-ces2018/

[85] http://archive.defense.gov/home/features/2015/0415_cyber-strategy/final_2015_dod_cyber_strategy_for_web.pdf

- The significant security and safety vulnerabilities presented to the organization by sharing or failing to protect this information;
- The increased mobility and interconnectedness of our personal and work-related data and devices;
- Poor cybersecurity practices with personal and work-related data and devices;
- Insufficient emphasis on enterprise-wide cybersecurity and cyberbiosecurity awareness raising, training, competency and compliance monitoring;
- Under estimation of the likelihood of a cyber intrusion;
- Failure to implement a cybersecurity plan that identifies and enforces proven cybersecurity practices including multi-factor authentication and rights management;
- Security vulnerabilities present in networked devices due to poor software design and/or a failure of manufacturers to issue patches for these flaws; and
- An inability or unwillingness of end users to proactively identify, consider and mitigate cyber vulnerabilities associated with networked equipment and systems.

Organizations must do more to acknowledge, mitigate and eliminate the cyber vulnerabilities present in the life science enterprise. Although the impact of a cyber penetration event can destroy an organization's reputation, be massively expensive, and present a threat to public health and national security, significant protection can be achieved with relative ease and small investment. The effective identification, elimination and mitigation of cyberbiosecurity vulnerabilities involves implementation of a management system approach to the application of cybersecurity principles and practices that culminate in the protection, monitoring, and hardening of all aspects of the biosecurity posture of the life science enterprise. To accomplish this, organizations must develop and implement a cybersecurity plan that inspires a culture of conscientious and continual awareness of potential cyberbiosecurity vulnerabilities associated with all communications, business sensitive information, data from networked laboratory devices and facility systems, and physical access to computer terminals. An effective cybersecurity plan will include deployment of a cyberbiosecurity handbook that sensitizes users to the implications of potential vulnerabilities, emphasizes the importance of security vigilance and drives the creation of an organizational culture that appreciates the value of enterprise data and the need to rigorously safeguard it. This does not mean that data and information cannot or should not be shared with collaborators, service contractors or other known parties who have a legitimate and authorized need for it. But it does mean that all staff members must develop a greater awareness of the potential sensitivity of this information and become proactive in its protection.

The cybersecurity plan must be grounded in clear policy stating that all enterprise data will be vigorously protected, limited in distribution, actively monitored for intrusion, theft and leakage, and will never be publicly available online. This means electronic communications, data streams and organizational information are encrypted at rest and in transit to prevent corruption or theft, are subject to secure cloud storage (or secure off-site storage) for redundancy, backup and resiliency to known and emerging threats, and are to be accessed only by known and trusted individuals utilizing cryptographically strong passwords and utilizing a properly implemented multi-factor authentication process. These solutions are commercially available off the shelf.

All organizational information must receive graded security protection through systematic classification (i.e., Public, Project Sensitive, Business Use Only, Restricted, Highly Confidential) and be subject to rigorous access control procedures including rights management to control the ability of individuals to view, edit, download, print and electronically distribute information both internally and externally. Employee access should be limited solely to that information necessary for the performance of their job.

While it's best to not work over unsecure public networks, this isn't always possible, which makes it wise to utilize a virtual private network to protect data communications.

Automated enterprise-wide IT security activities should include automated virus and malware scans for all emails and downloads, training of staff to recognize and report phishing scams, the monitoring of all staff activities on in-house electronic hardware, and monitoring of all data downloads and internet activities on the organization's networks to detect the insider threat.

Distribution of sensitive information to vendors, contractors, and service providers must be subject to non-disclosure agreements and must always be controlled, limited, and encrypted. Prior to distribution, the data owner should require and verify the parties possess appropriate policies, systems, technology and processes to similarly protect the owner's data. Otherwise, these parties represent an uncontrolled vulnerability to the enterprise. Under no circumstances should the parties be allowed to operate the organization's computer terminals, much less be provided direct access to the organization's network or data streams.

All staff members should be trained in the organization's cybersecurity practices with emphasis on the cyberbiosecurity vulnerabilities represented in the organization's data. Routinely, everyone should be assessed for their competence in these practices and actively monitored for compliance. Life science organizations should run anomaly detection software to identify and isolate threats as they emerge. They should also engage in penetration testing to ensure their systems can't be easily accessed from outside.

Future implementation of ML and AI tools to organizational data will present a significant improvement in cyberbiosecurity. These tools will be able to detect and respond to attempted cyber penetrations and thus prevent data theft and corruption from outside the organization. These same tools will enable organization leadership to improve compliance with security policy and practices to protect organizational data from the insider threat.

Aside from the implementation of cybersecurity practices, cybersecurity considerations must become a top priority before the deployment of any technology in the life science space. Scientists and other end users must recognize that every point of electronic interface presents a vulnerability. These individuals must not only train themselves to look at every piece of equipment to recognize and identify these points of vulnerability, they must also scrutinize all data and every process from a cybersecurity standpoint. Only from this vantage point can the user begin to thoroughly and comprehensively address the cyberbiosecurity risks in the life science enterprise. Additionally, hardware and software developers must proactively consider the security vulnerabilities of their products during the development process, not as an afterthought. These developers must make cybersecurity an immediate and fundamental component of their software and product design efforts.

CONCLUSIONS

The U.S. life science enterprise constitutes hundreds of thousands of biological laboratories[86]. Collectively, these laboratories comprise a significant portion of the U.S. gross domestic product (the bioeconomy). These life science laboratories possess cyberbiosecurity vulnerabilities associated with their networked hardware, devices and systems. These vulnerabilities pose an existential threat to individual organizations because exploitation of these vulnerabilities could jeopardize their reputation, the integrity and quality of research data, intellectual property, and biological products. Exploitation of these vulnerabilities could easily compromise the safety of building occupants, public health, the environment and national security.

Cyberbiosecurity vulnerabilities exist in large measure due to inadequate cybersecurity procedures, insufficient respect for the value of the organization's data, a failure to recognize the vulnerabilities revealed within the organization's sensitive business documents, and the failure of individuals to identify and address cybersecurity vulnerabilities associated with networked bench equipment, communication devices and facility systems. Equipment manufacturers and software developers shoulder responsibility, too. They fail to recognize, eliminate or mitigate the cybersecurity vulnerabilities in their products.

The digital transformation of today's laboratories into the smart labs of the future will be ushered in when virtual personal assistants are used to control networked equipment and systems. The application of artificial intelligence to virtual personal assistants networked across many organizations will assist decision making by senior leadership and institutional committees through the identification of cyberbiosecurity vulnerabilities and by providing recommendations for their elimination and/or mitigation. Wearables may be deployed in high risk laboratory environments to monitor human performance. Virtual reality may be widely adopted for training of laboratory staff, especially those performing work with animals and/or BSAT. Application of blockchain technology to create a verifiable and tamperproof record of every transaction made with laboratory data and digital genomic and protein sequences will guarantee the integrity of this information and provide an irrefutable means to authenticate and interrogate all manipulations. Smart labs will increase productivity and accelerate the adoption of additional disruptive technologies. As more networked devices and systems appear in the laboratory, the use of voice biometric authentication as part of multimodal biometric and multifactor authentication will significantly improve cybersecurity throughout life science organizations.

AUTHOR CONTRIBUTIONS

JR was the co-originator of cyberbiosafety concept and cyberbiorisk management concept, responsible for structure and content, responsible for considering, incorporating co-author contributions and suggested modifications, and responsible for final version. ND was the co-originator of cyberbiosafety concept and cyberbiorisk management concept, contribution to DNA synthesis hacking, voice biometric authentication, blockchain technology, and review and modifications to ensure overall content quality and readability.

FUNDING

Manuscript preparation and publication fees paid for by Inspirion Biosciences.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Robert J. Hawley, Lucia J. Reed, and Curtis Matthew Sharkey for their thoughtful review of this manuscript and helpful suggestions for improvement.

[86] Reed, personal data.

REFERENCES

Adam, L., Kozar, M., Letort, G., Mirat, O., Srivastiva, A., Stewart, T., et al. (2011). Strengths and limitations of the federal guidance on synthetic DNA. *Nat. Biotechol.* 29:208. doi: 10.1038/nbt.1802

Banerjee, A., Bandyopadhyay, T., and Acharya, P. (2013). Data analytics: Hyped up aspiration or potential? *Vikalpa* 38:1. doi: 10.1177/0256090920130401

Berger, K., and Roderick, J. (2014). *National and Transnational Security Implications of Big Data in the Life Sciences*. New York, NY: American Association for the Advancement of Science.

Burnette, R. N., Hess, J. E., Kozlovac, J. P., and Richmond, J. Y. (2013a). "Defining biosecurity and related concepts" in *Biosecurity – Understanding, Assessing, and Preventing the Threat*, ed R. Burnette (Hoboken, NJ: John Wiley and Sons, Inc.), 3–16.

Burnette, R. N., Reed, J. C., and Delarosa, P. (2013b). "The future of biosecurity: a global context" in *Biosecurity – Understanding, Assessing, and Preventing the Threat*, ed R. Burnette (Hoboken, NJ: John Wiley and Sons, Inc.), 259–269.

Department of Homeland Security (2016). *Healthcare and Public Health Sector-Specific Plan*. Department of Homeland Security.

Department of the Army (2008). *Army Regulation 50-1 Biological Surety*. Department of the Army.

Gudivada, V. N., Irfan, M. T., Fathi, E., and Rao, D. L. (2016). "Chapter 5: Cognitive analytics: going beyond big data analytics and machine learning" in *Handbook* *of Statistics*, eds V. N. Gudivada, V. V. Raghavan, V. Govindaraju, and C. R. Rao (North Holland: Elsevier), 169–205.

Hao, W., and Chuen-Tsai, S. (2011). "Game reward systems: gaming experiences and social meaning," in *Proceedings of the 2011 DiGRA International Conference: Think Design Play. 5th International Conference of the Digital Research Association: Think Design Play, DiGRA 2011, Vol. 6* (Utrecht). Available online at: www.digra.org/wpcontent/uploads/digitallibrary/11310.20247.pdf

Kumar, K., and Farik, M. (2016). A review of multimodal biometric authentication systems. *Int. J. Sci. Technol. Res*. 5, 5–9.

Kwok, R. (2018). How to pick an electronic laboratory notebook. *Nature* 560:269. doi: 10.1038/d41586-018-05895-3

McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (2006). A proposal for the Dartmouth summer conference on artificial intelligence. *AI Magazine* 27, 12–14

Meyerson, L. A., and Reaser, J. K. (2002). A unified definition of biosecurity. *Science* 295:44. doi: 10.1126/science.295.5552.44a

Murch, R. S., So, W. K., Buchholz, W. G., Raman, S., and Peccoud, J. (2018). Cyberbiosecurity: an emerging new discipline to help safeguard the bioeconomy. *Front. Bioeng. Biotechnol.* 6:39. doi: 10.3389/fbioe.2018.00039

National Association of County and City Health Officials (2016). *2016 National Profile of Local Health Departments*. National Association of County and City Health Officials.

National Institute of Standards and Technology (2018). *Framework for Improving Critical Infrastructure Cybersecurity, Version 1.1.*

Peccoud, J., Gallegos, J. E., Murch, R., Buchholz, W. G., and Raman, S. (2017). Cyberbiosecurity: from naïve trust to risk awareness. *Trends Biotechnol.* 36:12. doi: 10.1016/j.tibtech.2017.10.012

Reed, J. C., Heckert, R., Delarosa, P., and Ettenger, V. (2013). "Utilizing biosecurity principles to combat naturally occurring epidemics" in *Biosecurity – Understanding, Assessing, and Preventing the Threat*, ed R. Burnette (Hoboken, NJ: John Wiley and Sons, Inc.), 167–183.

Reed, J. C., and Sharpe, D. C. (2013). "Operational elements of biosecurity" in *Biosecurity – Understanding, Assessing, and Preventing the Threat*, ed R. Burnette (Hoboken, NJ: John Wiley and Sons, Inc.), 71–88.

Robson, K., Plangger, K., Kietzmann, J., McCarthy, I., and Pitt, L. (2015). Is it all a game? Understanding the principles of gamification. *Business Horizons*. 58:411. doi: 10.1016/j.bushor.2015.03.006

Ruan, S., Wobbrock, J. O., Liou, K., Ng, A., and Landay, J. A. (2017). "Comparing speech and keyboard text entry for short messages in two languages on touchscreen phones," in *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (New York, NY). doi: 10.1016/bs.host.2016.07.010

Shannon, C. E., and McCarthy, J. (eds.). (1956). *Automata Studies*. Princeton, NJ: Princeton University Press.

Shubhendu, S., and Vijay, J. (2013). Applicability of artificial intelligence in different fields of life. *Intl. J. Sci. Eng. Res.* 1, 2347–3878.

Singh, A., Thakur, N., and Sharma, A. (2016). "A review of supervised machine learning algorithms," in *3rd International Conference on Computing for Sustainable Global Development (INDIACom)* (New Delhi).

Skvorc, C., and Wilson, D. E. (2011). Developing a behavioral health screening program for BSL-4 laboratory workers at the National Institutes of Health. *Biosec. Bioterr.* 9:23. doi: 10.1089/bsp.2010.0048

The National Academies of Sciences Engineering and Medicine (2015). *Meeting Recap, Safeguarding the Bioeconomy: Applications and Implications of Emerging Sciences*. Washington, DC: Organized by Board on Chemical Sciences and Technology.

Turing, A. M. (1950). Computing machinery and Intelligence. *Mind* 49:433. doi: 10.1093/mind/LIX.236.433

United States Government Accountability Office (2012). *Medical Devices – FDA Should Expand its Consideration of Information Security for certain Types of Devices*. United States Government Accountability Office

World Health Organization (2004). *Laboratory Biosafety Manual, 3rd Edn*. Geneva: World Health Organization.

World Health Organization (2006). *Biorisk Management – Laboratory Biosecurity Guidance*. Geneva: World Health Organization.

Yan, J. (2017). Percentage of Americans taking antidepressants climbs. *Psychiatric News* 52:1. doi: 10.1176/appi.pn.2017.pp9b2

Conflict of Interest Statement: The authors declare they were employees of Inspirion Biosciences during the period in which this manuscript was prepared and that this manuscript was prepared in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read
for greatest visibility
and readership

FAST PUBLICATION
Around 90 days
from submission
to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative,
and constructive
peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers
acknowledged by name
on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data
and methods to enhance
research reproducibility

DIGITAL PUBLISHING
Articles designed
for optimal readership
across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics
track visibility across
digital media

EXTENSIVE PROMOTION
Marketing
and promotion
of impactful research

LOOP RESEARCH NETWORK
Our network
increases your
article's readership