Deploying Digital Detection of Dangerous DNA

Hanna Pálya[1*], Oscar Delaney[2*]
[1]The University of St. Andrews, United Kingdom
[2]The University of Queensland, Australia
*All authors contributed equally
https://doi.org/10.38126/JSPG210306
Corresponding author: hannapalya@gmail.com
Keywords: biosecurity; biotechnology; technology policy; biosafety

Executive Summary: Ready access to synthetic DNA has enabled vast progress in health and biotechnology in recent decades. However, it also makes it easier for malicious actors to develop biological weapons that threaten public health. Thus, it is crucial to regulate who can order synthetic DNA of pathogens from commercial providers. Some synthetic DNA providers screen customers and ordered sequences; however, there is no federal legislation addressing the responsibilities of these companies. Voluntary compliance with the screening Guidance issued by the US Department of Health and Human Services (HHS) is not sustainable, as it disadvantages companies who screen against those who do not. Here, we consider two policy solutions. Option 1 is to mandate minimum screening standards for all US-based synthetic DNA providers. This regulatory burden may weaken the US DNA synthesis sector, so we recommend Option 2, which requires government-funded researchers to only use synthetic DNA from accredited screening-compliant providers.

I. The biosecurity threat of synthetic DNA

The ongoing COVID-19 pandemic has highlighted the grave danger posed to humanity by the spread of novel pathogens. As synthetic biology and digital technology progress, humanity is becoming better equipped to prevent, detect, and respond to natural pandemics, but simultaneously, the risks from human-caused pandemics are increasing (Madhav et al. 2017, 317-321). This is partly due to the development of user-friendly synthetic biology tools that could grant increasingly many people unaffiliated with academic or industrial research groups the ability to create and release deadly pathogens (Dev 2018, 3-17). A key concern is that a bioterrorist group or another malicious actor would order synthetic DNA from a commercial provider and use this to construct a bioweapon.[1]

There are no publicly known cases of synthetic DNA being used maliciously; however, the history of bioterrorism suggests this is a serious concern. The Japanese apocalyptic cult Aum Shinrikyo, for instance, ran a multimillion-dollar secret bioweapons program involving anthrax, Ebola virus and botulinum toxin between 1990 and 1995, but failed to deploy functional and dangerous strains of these organisms (Danzig et al. 2012, 18-28). If a similar group were active today, the path to causing severe harm would be easier.

While the annual probability of a bioterrorist causing a terrible pandemic is small, the potentially catastrophic nature of the risk means it deserves serious public policy attention. Indeed, deliberately released pathogens could be far worse than naturally arising ones, as lethality and transmissibility may be artificially enhanced.

Along with the risk from deliberate attacks like the 2001 anthrax letters (Jernigan et al. 2001, 934), accidental lab leaks such as the one resulting in the 2007 foot-and-mouth disease outbreak (Anderson 2008, 8-13) pose an increased threat if researchers are using dangerous DNA. Even if any adverse incidents remain limited in scope, they could plausibly turn public opinion against synthetic biology. This potential overreaction would hinder

[1]We deliberately do not provide any details on these techniques, in accordance with infohazard best practices.

the development and uptake of new technological solutions to public health issues.

In the coming decades we will likely discover more pathogens that pose even graver dangers than currently endemic diseases, so safeguarding all synthetic DNA before then is crucial.[2]

Synthetic DNA provides many benefits, especially in the research and commercialization of medical technology, food science, and biomanufacturing, so even pandemic risks should not necessarily be seen as sufficient reason to shut down this sector (Engineering Biology Leadership Council 2021, 6-8). The US leads the world in synthetic biology, accounting for 39.8% of the global market, and this is where many of the best and biggest DNA synthesis companies are located, so it is especially important that biotechnology is used safely and effectively here (Global Industry Analysts 2022).

II. Safeguarding DNA synthesis

To assess the efficacy of safety measures, we cannot measure reductions in malicious use of DNA synthesis, as there is a baseline of zero. Instead, we must use some combination of theoretical analysis and 'red-teaming,' where safe actors deliberately make dangerous orders, to verify DNA companies have adequate security measures to catch the danger. Twist Bioscience, a large and well-respected DNA synthesis company, red-teamed their screening systems and recommended other companies do likewise (Diggans and Leproust 2019, 2).

Many academics are working on developing and promoting better tools for DNA synthesis screening[3] and some policy groups, notably the Nuclear Threat Initiative in collaboration with the World Economic Forum, are lobbying for improved government regulation (Nuclear Threat Initiative and World Economic Forum 2020). There is comparatively little opposition to reform, as no one stands to gain from bioweapons being easily available. At worst, companies are reluctant to spend extra money on safety protocols; they are not opposed in principle. This lack of concerted opposition means regulatory reform is more achievable than in many domains.

One way to avert disaster is to very strictly police access to the genetic sequence data of dangerous pathogens. It is too late to implement this for already known pathogens, as their genomes are freely available on the internet, including 23,816 viral genomes on the National Centre for Biotechnology Information (NCBI) database alone (Vinatzer et al. 2019, 3).[4] However, it is an important and urgent recommendation for pathogens we are yet to discover, given that the number of genomes listed on the NCBI database has been doubling on average once every eighteen months (NCBI 2022). The research community could maintain almost all the value of open science data-sharing practices (Hetu, Koutouki, and Joly 2019, 1-2) while withholding the genomes of the few very dangerous pathogens.

Another approach is to tightly regulate who has access to synthetic DNA and only allow trusted researchers in academic labs and biotechnology companies to use this technology (Esvelt 2018, 2-5). This can be achieved through thorough customer screening to check that the customer is a legitimate researcher, similar to a background check. While customer screening of synthetic DNA orders should be done to some extent, taking this too far will pose a bureaucratic burden that will slow down science and innovation, including medical research, and thus be counterproductive. Moreover, there is some danger that a legitimate researcher at a known institution will turn out to be a malicious actor, so customer screening is insufficient.

The most implementable and effective screening solution centers on digital technology to computationally screen all incoming synthetic DNA orders to determine whether they could be used for a bioweapon. The simplest instantiation of this is to check whether an order is a close match to a gene from any pathogen of concern. However, some benign sequences are very similar across pathogenic and non-pathogenic organisms, and these sequences being falsely flagged as dangerous could hamper

[2] Of particular concern is USAID's DEEP VZN program which seeks to discover novel zoonotic pathogens, publish their genomes, and characterize their pandemic potential (Sandbrink et al. 2022).

[3] In the authors' opinion, the most important technical work is being done by Secure DNA (part of Kevin Esvelt's lab at MIT), Todd Treangen's lab at Rice University, and Battele's Threatseq project, though many of the people listed in the references are doing valuable work.

[4] Most of these are not pathogenic to humans, though some are.

valuable research (Diggans and Leproust 2019, 2). More recently, software has been developed to detect how similar a given order is to sequences with dangerous functions, and hence whether to approve the order (Balaji et al. 2022, 1-3). A particularly promising approach involves cryptographically encoding a database obtained through international consensus that contains all known dangerous sequences and checking every order for matches against this database (Gretton et al. n.d., 2-6). Encryption allows all companies to use the database, while no company can access the unencrypted data of the dangerous sequences.

Currently, many companies do screen using the tools described above, notably those in the International Gene Synthesis Consortium (IGSC), an industry body set up in 2009 for companies to coordinate screening best practices. However, no government currently mandates that companies screen all orders, and the UN and other intergovernmental bodies have also not taken charge on this issue. Thus, it remains straightforward for malicious actors to order from companies they know do not screen. One of the key bottlenecks for achieving a world safer from engineered pandemics is to enact policy and governance interventions to make more companies meet minimum screening standards.

III. Policy options

i. Option 1: Mandating minimum screening standards

Currently, the Department of Health and Human Services (2022, 25496) states DNA synthesis companies should be "verifying the legitimacy of customers" and checking for "sequences derived from or encoding select agents and toxins." However, there is no legal obligation for companies to comply with this Guidance. As of today, eleven US DNA synthesis companies are members of the IGSC and hence perform some screening, while at least twelve others show no evidence of performing screening.[5]

One policy option is for Congress to enact legislation requiring that all synthetic DNA providers must meet minimum customer and sequence screening standards, similar to those used by the IGSC (International Gene Synthesis Consortium 2017).[6]

Advantages

This is the clearest route to achieving universal screening in the US, and it is relatively simple to administer as there would be a defined benchmark specific companies must meet. This option has many parallels in other industries; for instance, pharmacists must check whether each order is dangerous and if so only allow a customer to complete the purchase if they have a prescription (Code of Federal Regulations, n.d.). Because the US is a global leader in DNA synthesis, the successful implementation of this law could set a powerful precedent for other countries to follow. Moreover, regardless of other countries passing similar legislation, foreign DNA synthesis companies would need to comply with US standards to be allowed to export to the large US market.

Disadvantages

Overly onerous DNA synthesis screening regulation would introduce significant new costs of doing business as a DNA provider in the US, and hence push companies overseas. This would hinder the US economy, especially research and development, and harm synthesis company employees. Moreover, it may even be counterproductive from a biosecurity perspective if excessive regulation drives greater market share to countries with worse safety and security infrastructure. Thus, to be worthwhile, this ban on unsafe synthesis companies must be introduced gradually, while assisting companies to develop compliant screening systems with minimal additional cost.

Enforcing this new regulation would require government agencies to monitor the screening practices of companies, perhaps including red-teaming them. Further, the minimum standards would need to be regularly revisited to ensure they are appropriate in light of ongoing rapid

[5]The IGSC website lists member companies at https://genesynthesisconsortium.org/. We will not disclose here the names of the companies that appear not to screen, lest this information help malicious actors. We are happy to supply our database of DNA providers categorized by screening status and country to trusted researchers upon request.

[6] The HHS Guidance could also be used as a model, though the IGSC standards are preferable as they provide more flexibility to implement screening with different techniques, while the HHS Guidance is more rigid and quickly outdated.

technological change. All this will require additional government expenditure without increasing taxation, so will not be revenue-neutral.

ii. Option 2: Requiring government-funded research to only use synthetic DNA from accredited providers

The HHS Guidance is directed towards synthetic DNA providers, outlining screening standards that they should follow, which pose extra costs and might deter some customers. While DNA providers have a great responsibility to ensure their products are safe, they should not be the sole actors bearing the burden of screening. Without customer compliance, the efforts of synthesis companies who screen are futile, given that researchers who do not want their orders to be screened could order from companies that do not screen. Indeed, if an order is rejected on biosecurity grounds, there is nothing to stop this customer from simply re-ordering the same sequences from a non-screening company.

We propose that Congress enact legislation requiring that all government-funded researchers only use synthetic DNA from companies that meet minimum customer and sequence screening standards, similar to those used by the IGSC (International Gene Synthesis Consortium 2017).

Advantages

The federal government provides massive funding for research and development, $138B in 2020 alone, which was 19% of all R&D funding in the US (Boroush 2022).[7] Much of this is spent on life sciences research, including synthetic biology. Thus, ensuring that none of this research funding is used to purchase DNA from non-screening synthesis companies will create significant extra demand, and thus revenue, for compliant DNA providers. Consequently, it would also motivate non-screening companies to start screening to avoid losing federally funded customers. Further, this policy could raise awareness of the importance of biosecurity among researchers, leading to better security standards and norms.

Disadvantages

Government-funded researchers are unlikely to be covertly pursuing bioweapons, so arguably this policy is not targeted at the gravest risk. Malicious actors are less likely to be receiving government funding and could therefore easily keep ordering from non-screening companies with private funding. However, as stated above, this policy would indirectly incentivize companies to start screening, eventually eliminating the non-screening companies that malicious actors could order from.

Another difficulty is that federal legislation is hard to enact. The State of California tried to mandate that the University of California (UC) and California State University (CSU) only use screened DNA, however, the legislation was vetoed by the Governor and later softened to only require CSU to provide guidance to their researchers and request UC to do the same (California State Legislature 2022). If a stronger version could not be enacted with bipartisan State legislature support, it is hard to imagine how it could work on a federal level. However, even if legislation proves to be too hard, it is still worthwhile to at least provide thorough guidance and recommendations for researchers using government funding, given the significant efforts of many DNA providers to follow the voluntary HHS Guidance.

IV. Policy recommendation

We recommend Option 2, to require government-funded researchers to only use synthetic DNA from companies that meet minimum customer and sequence screening standards. To implement this policy, the federal government would maintain a list of domestic and international companies that have passed all necessary checks and been accredited as safe providers. Researchers would then need to acknowledge in their grant application that they are legally required to only use DNA from accredited companies. Should there be concerns about researchers reneging on this commitment, a small enforcement team conducting spot checks would likely prove sufficient to keep researchers compliant. Enacting this policy would improve DNA synthesis security by encouraging more companies to screen and would not substantially hinder synthetic biology research. Thus, this proposal would reduce public health risks and ensure that we can continue to safely harness the power of biotechnology.

[7] The remaining R&D funding came from corporations, universities, philanthropists and state governments.

References

Anderson, Iain. 2008. Foot and Mouth Disease 2007: A Review and Lessons Learned. (House of Commons). https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/250363/0312.pdf.

Balaji, Advait, Bryce Kille, Anthony D. Kappell, Gene D. Godbold, Madeline Diep, R. A. Leo Elworth, Zhiqin Qian, Dreycey Albin, Daniel J. Nasko, Nidhi Shah, Mihai Pop, Santiago Segarra, Krista L. Ternus, and Todd J. Treangen. 2022. "SeqScreen: accurate and sensitive functional screening of pathogenic sequences via ensemble learning." Genome Biology 23 (1): 133. https://doi.org/10.1186/s13059-022-02695-x.

Boroush, Mark. 2022. U.S. R&D Increased by $62 Billion in 2019 to $667 Billion; Estimate for 2020 Indicates a Further Rise to $708 Billion. (National Center for Science and Engineering Statistics). https://ncses.nsf.gov/pubs/nsf22330.

California State Legislature. 2022. AB-1963: California State University and University of California: gene synthesis providers. https://leginfo.legislature.ca.gov/faces/billStatusClient.xhtml?bill_id=202120220AB1963.

Code of Federal Regulations. Title 21, Chapter II, Part 1306. National Archives. https://www.ecfr.gov/current/title-21/chapter-II/part-1306.

Danzig, Richard, Marc Sageman, Terrance Leighton, Lloyd Hough, Hidemi Yuki, Rui Kotani, and Zachary M. Hosford. 2012. Aum Shinrikyo: Insights Into How Terrorists Develop Biological and Chemical Weapons. Center for a New American Security. http://www.jstor.org/stable/resrep06323.

Department of Health and Human Services. 2022. Screening Framework Guidance for Providers and Users of Synthetic Oligonucleotides https://www.federalregister.gov/documents/2022/04/29/2022-09210/screening-framework-guidance-for-providers-and-users-of-synthetic-oligonucleotides.

Dev, Ishaan. 2018. Democratizing Synthetic Biology: Balancing Biosecurity, Biosafety, and Citizen Science. Washington Internships for Students of Engineering. https://wise-intern.org/wp-content/uploads/2019/04/WISE_AIChE-Final-Draft_Ishaan-Dev.pdf.

Diggans, James, and Emily Leproust. 2019. "Next Steps for Access to Safe, Secure DNA Synthesis." Frontiers in Bioengineering and Biotechnology 7. https://doi.org/10.3389/fbioe.2019.00086.

Engineering Biology Leadership Council. 2021. Engineering Biology for the UK - A Resource to help Build Back Better. https://ktn-uk.org/wp-content/uploads/2021/07/EBLC-Building-back-better-with-Engineering-Biology_upload.pdf.

Esvelt, Kevin M. 2018. "Inoculating science against potential pandemics and information hazards." PLOS Pathogens 14 (10): e1007286. https://doi.org/10.1371/journal.ppat.1007286.

Global Industry Analysts. 2022. A $28.8 Billion Global Opportunity for Synthetic Biology by 2026. Cision PR Newswire. https://www.prnewswire.com/news-releases/a-28-8-billion-global-opportunity-for-synthetic-biology-by-2026---new-research-from-strategyr-301493554.html.

Gretton, Dana, Erika A. DeBenedictis, Andrew B. Liu, Andrew C. Yao, and Kevin M. Esvelt. n.d. "Random adversarial threshold search enables specific, secure, and automated DNA synthesis screening." preprint. https://www.securedna.org/manuscript/Random_Adversarial_Threshold_Screening.

Hetu, Martin, Konstantia Koutouki, and Yann Joly. 2019. "Genomics for All: International Open Science Genomics Projects and Capacity Building in the Developing World." Frontiers in Genetics 10. https://doi.org/10.3389/fgene.2019.00095.

International Gene Synthesis Consortium. 2017. Harmonized Screening Protocol v2.0: Gene Sequence & Customer Screening to Promote Biosecurity https://genesynthesisconsortium.org/wp-content/uploads/IGSCHarmonizedProtocol11-21-17.pdf

Jernigan, John A., David S. Stephens, David A. Ashford, Carlos Omenaca, Martin S. Topiel, Mark Galbraith, Michael Tapper, Tamara L. Fisk, Sherif Zaki, Tanja Popovic, Richard F. Meyer, Conrad P. Quinn, Scott A. Harper, Scott K. Fridkin, James J. Sejvar, Colin W. Shepard, Michelle McConnell, Jeannette Guarner, Wun-Ju Shieh, Jean M. Malecki, Julie L. Gerberding, James M. Hughes, and Bradley A. Perkins. 2001. "Bioterrorism-Related Inhalational Anthrax: The First 10 Cases Reported in the United States." Emerging Infectious Disease journal 7 (6): 933. https://doi.org/10.3201/eid0706.010604.

Madhav, Nita, Ben Oppenheim, Mark Gallivan, Prime Mulembakani, Edward Rubin, and Nathan Wolfe. 2017. "Pandemics: Risks, Impacts, and Mitigation." In Disease Control Priorities, Third Edition: Volume 9. Improving Health and Reducing Poverty, edited by Dean T. Jamison, Hellen Gelband, Susan Horton, Prabhat Jha, Ramanan Laxminarayan, Charles N. Mock and Rachel Nugent. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-0527-1_ch17

National Centre for Biotechnology Information (NCBI). 2022. "GenBank and WGS Statistics." National Library of Medicine. https://www.ncbi.nlm.nih.gov/genbank/statistics/

Nuclear Threat Initiative, and World Economic Forum. 2020. Biosecurity Innovation and Risk Reduction: A Global Framework for Accessible, Safe and Secure DNA Synthesis. (Geneva). https://www3.weforum.org/docs/WEF_Biosecurity_Innovation_Risk_Reduction.pdf.

Sandbrink, Jonas B., Janvi Ahuja, Jacob L. Swett, Gregory D. Koblentz, and Claire J. Standley. 2022. "Mitigating Biosecurity Challenges of Wildlife Virus Discovery." SSRN. https://doi.org/10.2139/ssrn.4035760.

Vinatzer, Boris A., Lenwood S. Heath, Hussain M. J. Almohri, Michael J. Stulberg, Christopher Lowe, and Song Li. 2019. "Cyberbiosecurity Challenges of Pathogen Genome Databases." Frontiers in Bioengineering and Biotechnology 7. https://doi.org/10.3389/fbioe.2019.00106.

Hanna Pálya is a biosecurity researcher at the Cambridge Existential Risk Initiative, with a research focus on public health and catastrophic biological risks. She is a final-year undergraduate at the University of St Andrews, studying biology and philosophy. Previously, she worked for a Hungarian climate policy think tank, researching climate philosophy; and interned at the Cambridge Infectious Diseases Centre.

Oscar Delaney is a biosecurity researcher at the Cambridge Existential Risk Initiative, with a research focus on biotechnology policy and regulation. He holds a Bachelor of Science and Bachelor of Arts from the University of Queensland, with majors in genetics, mathematics and philosophy. He previously interned as a quantitative trader at a finance firm, and as a research assistant to several academics.