COMMENTARY

Beyond Biosecurity by Taxonomic Lists: Lessons, Challenges, and Opportunities

Piers Millett, Tessa Alexanian, Kathryn R. Brink, Sarah R. Carter, James Diggans, Megan J. Palmer, Ryan Ritterson, Jonas B. Sandbrink, and Nicole E. Wheeler

INTRODUCTION

Taxonomic lists—that is, lists that give a set of species (eg, *Yersinia pestis*) or broader taxonomic groups (eg, SARS-related coronavirus) to be regulated—are a common tool used within biosecurity regimes for controlling access to pathogenic organisms and toxins. However, these lists are both too specific and too ambiguous for many of the uses to which they are applied. Their use is based on an assumption that taxonomy can be directly linked to biological function; however, other approaches, such as sequence lists or tools that predict function, may be more suitable to capture biological functions that cause concern. We are not the first to notice the problems with taxonomic lists; their shortcomings have been highlighted for over a decade.[1-4]

We now have a wealth of experience in using taxonomic lists. Advances in biological engineering have both further strained taxonomy-based control systems and offered new alternatives to them. For example, in their 2010 report,[3] the US National Academy of Sciences noted that a system for predicting risk—a combination of the probability of harm occurring and the severity (consequences) of that harm, if it were to occur[5]—from genetic data was not possible then or in the "usefully near future." We believe things have changed. Two US Intelligence Advanced Research Projects Activity programs—Functional Genomic and Computational Assessment of Threats (FunGCAT)[6] and Finding Engineering-Linked Indicators[7]—have provided a drive toward functional prediction based on sequence, which is being applied directly to commercial DNA synthesis screening.[8] Recent advances in protein structure prediction, combined with advances in generative models and high-throughput screening, promise to usher in a new era of protein design and formally remove the previously reliable assumption of a link between taxonomic origin and biological function.[9-11]

Piers Millett, PhD, is Executive Director, International Biosecurity and Biosafety Initiative for Science, Washington, DC. Tessa Alexanian is Safety and Security Program Officer, iGEM Foundation, Paris, France. Kathryn R. Brink, PhD, is a Postdoctoral Fellow, Center for International Security and Cooperation, and Megan J. Palmer, PhD, is Executive Director of Bio Policy & Leadership Initiatives and an Adjunct Professor, Department of Bioengineering; all at Stanford University, Stanford, CA. Sarah R. Carter, PhD, is Principal, Science Policy Consulting LLC, Arlington, VA. James Diggans, PhD, is Head of Biosecurity, Twist Bioscience, San Francisco, CA. Ryan Ritterson, PhD, is Executive Vice President of Research, Gryphon Scientific LLC, Takoma Park, MD. Jonas B. Sandbrink is a Doctoral Researcher, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom. Nicole E. Wheeler, PhD, is a Turing Fellow, Institute of Microbiology and Infection, University of Birmingham, Birmingham, United Kingdom.

Recent policies and guidance have attempted to encode broad principles in place of taxonomic lists. At the international level, the latest World Health Organization *Laboratory Biosafety Manual*[12] focuses on broad risk assessment* over taxonomic classification.[13] However, at the national or institutional level, lists of species remain a common starting point for risk management.[†14-16] The US *Recommended Policy Guidance for Departmental Development of Review Mechanisms for Potential Pandemic Pathogen Care and Oversight*[17] covers any research that is "reasonably anticipated to create, transfer or use potential pandemic pathogens resulting from the enhancement of a pathogen's transmissibility and/or virulence in humans."[18] Several challenges with these policy experiments have already emerged, such as ambiguous scope,[19] disagreements on what their text means in practice (eg, the term "endow or enhance pathogenicity" in export control rules[20]), a need for greater transparency in decisionmaking,[21] and guidance on responsible communication.[22]

We do not advocate abandoning taxonomic lists entirely. Taxonomic lists are an excellent tool for controlling access to whole organisms, which is necessary for export controls[20,23] and when possession of certain agents has been criminalized.[24,25] They can also have implications for the practical implementation of other policies where the best control point is a whole organism. However, modern biotechnology increasingly makes use of parts of organisms to confer new traits to species, sometimes making the origin species of a single part a poor (or even misleading) proxy for risk assessment. New, complementary tools are needed.

We believe that taxonomic lists become less useful when used beyond their intended context. Although they were originally created to control physical access to organisms, taxonomic lists have been integrated into broader biosecurity policies and risk management processes. For example, gene synthesis screening routinely uses taxonomic classification as a core element of risk assessment. Commercial providers of DNA synthesis have noted the challenges of implementing taxonomy-based regulations for sequence-based products.[26,27] Ascertaining whether a given order matches something to be controlled is difficult even for the most qualified companies. Taxonomic lists have also been used as a tool to narrow oversight of relevant biological research. For instance, although US dual use research of concern policies are intended to consider more indirect scenarios of misuse, including generation of "knowledge, information, products, or technologies,"[28] the

scope of the oversight is restricted to specific types of experiments, with a specific set of pathogens on a taxonomic list. This scope may not be a good fit for the work that warrants oversight; more than two-thirds of life scientists who declared that their work featured "dual-use potential" in a 2009 survey would not fall into the policy's scope.[29]

In this commentary, we highlight some of the challenges encountered with taxonomic lists in practice. We explore opportunities to make them more effective. We also stress the importance, and attempt an initial discussion, of how we might move beyond taxonomic lists to meet an evolving landscape of safety and security needs.

A More Structured Approach to Designing Biosecurity Regimes

Too often, we use taxonomic lists because policymakers and regulators are familiar with them, but they need not be the default option. Not all policy purposes will require taxonomic lists, and we argue for a more deliberative approach (Figure).

The first step in developing a biosecurity regime should be a thorough understanding of its goals. What does the regime intend to accomplish? Is the goal to control *access* to particular materials, such as those likely to harm human health? Alternatively, the goal might be to control certain criminal *uses* of materials. Presumably, in all cases, the intent is also to avoid unintentionally impeding or unnecessarily restricting open, reproducible, and beneficial science and engineering. A regime that can articulate the ultimate goals and their underlying logic and principles is more likely to be able to be updated as assumptions are checked and revisited.

A second step is to translate the goals of the regime into biological functions to be controlled. A simple example is the goal of controlling pathogens that cause certain diseases, where the function to control is causing the disease. A more complex goal is controlling pathogens that meet thresholds for environmental stability, transmissibility, or virulence or for which vaccines and therapeutics do not exist. This goal recognizes the possibility of novel pathogens emerging that have similar biological properties to those known today. An even broader goal is to manage potential harms other than disease, such as those from gene drives (where the function would be to alter the inheritance of a certain trait) or antimateriel agents (where the function might be, for example, breaking down electronics). We assume that any regime's goal will focus on a subset of biological functions that pose a high risk of harm. This exempts other biological functions from control, thereby minimizing any undue impact on biological research and development.

The final phase is to identify which tools are best suited to the functions to be controlled. For example, a taxonomic list would be suited to a regime intended to control access to certain pathogens and to make their unauthorized possession a crime. By contrast, a regime aiming to control access to certain biological functions may need different

*Risk assessment: A systematic process—quantitative or qualitative—of gathering information and evaluating the nature, probability, and magnitude of potential harms and determining the appropriate control measures to minimize or otherwise mitigate the risks.[5]

†Risk management: The quantitative or qualitative forecasting and evaluation of the probability of harm occurring and subsequent consequences (risk assessment) together with the identification and implementation of technologies, measures, or practices to avoid or minimize their likelihood or impact.[5]
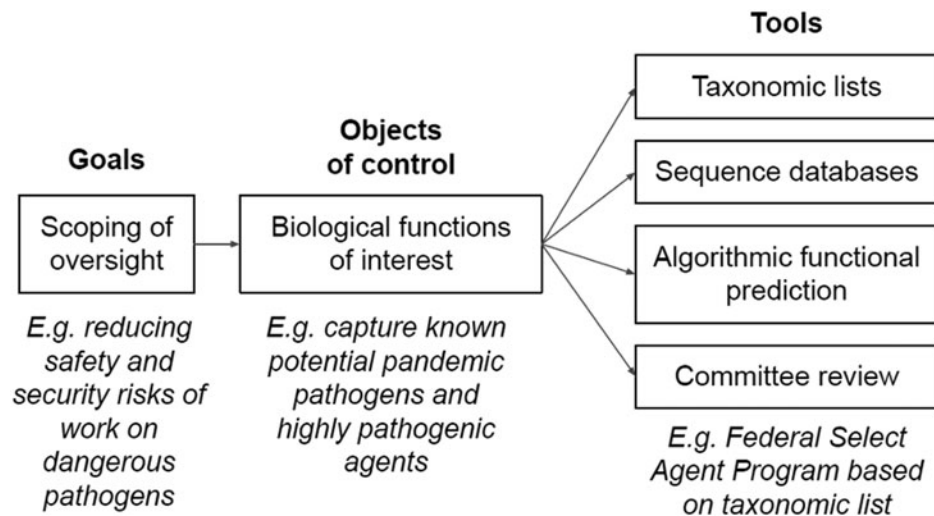
Figure. Tools for effective biosecurity policies. The goal and scope of a given regulation are a political decision. Once a scope has been set, regulators need to define the target objects with what functions should be targeted. What research objects feature these functions and are thus regulated can be defined using different tools.

tools, such as sequence databases. Recent efforts, such as FunGCAT,[6] Secure DNA,[30] SeqScreen,[31] and the Common Mechanism for DNA Synthesis Screening,[32] demonstrate the feasibility of compiling such databases. Emerging algorithms to predict function from sequence may be adopted when it becomes necessary to control access to genetic elements from novel or engineered sequences. We have identified a series of general rules when using any type of lists in these tools (Box).

Biosecurity regimes developed for different purposes in different locations by different institutions or by different communities will need different types of tools. Some tools may already exist (eg, taxonomic lists). Other tools (eg, sequence databases) may have been created for other purposes and may need to be reconfigured or curated for use in biosecurity regimes.[27] Some tools needed by biosecurity regimes are not yet fully developed (eg, algorithms to identify harmful functions from primary DNA or protein sequences). When developing tools, it is vital to take into account the need to avoid unduly impeding biological research and development and to manage any information hazards generated.

IMPROVING TAXONOMIC LISTS

There are at least 5 ways that taxonomic lists might be improved to increase and potentially extend their utility. Some of these would help address how we capture pathogenicity and virulence. Others will help address harms beyond traditional definitions of disease to address environmental disruption of antimaterial applications.

Conduct More Regular and Structured Updates

Taxonomic lists must be regularly updated to respond to the emergence of novel pathogens, scientific discoveries, advances in engineering capabilities, and geopolitical factors that may impact threat models for misuse. Yet, in practice, policy updates can be slow to materialize. For example, a version of *Bacillus cereus* that features 2 plasmids similar to those determining the pathogenicity of *Bacillus anthracis* and causes anthrax-like disease was identified in 2010.[33,34] Although most *B cereus* strains are not of concern, this strain, *B cereus serovar anthracis*, includes the concerning spore-forming and toxin-producing properties of *B anthracis*. Six years after this discovery, in 2016, the US Select Agents and Toxins List[25] expanded to also capture "anthrax-like *B cereus*."[35] Many other taxonomy-based control lists have still not been updated to capture this pathogen, highlighting the importance of approaches and regimes that can be updated in a timely manner.[20,36]

Include Functional Equivalence

Those responsible for maintaining taxonomic-based lists should commit to including functionally equivalent pathogens. This should be accompanied by annotating in their lists the criteria for the inclusion or exclusion of pathogens as well as the policy goals. This will make it easier to identify other pathogens that meet the same criteria based on functional equivalence. *B cereus serovar anthracis* is also a useful example of functional equivalency.

Including pathogens via functional equivalency could pose certain challenges. Including functional equivalence should not be used beyond control of whole organisms because it can be difficult, if not impossible, to link individual genetic elements as necessary or sufficiently causative of observed functional equivalence. In addition, including functional equivalence will require careful review for information hazards, for example, avoiding the publication of novel threat applications when explaining why a function is being included.

Box. General Guidelines for the Use of Lists Within Biosecurity Regimes

1. **Determine whether a list of specific biological materials, a list of functions, or a set of broad principles is more appropriate for a given context.** Lists of biological organisms or sequences are useful tools for contexts in which narrow and specific reference to controlled agents is required, for instance for law enforcement. However, such lists only capture known concerns. Thus, in settings where novel risks are meant to be anticipated, lists of functions or sets of broader principles rather than lists of specific biological materials may be better suited to guide decisionmaking.

2. **Clearly state whether a list is meant to be illustrative or exhaustive.** Better contexts around lists and their underlying principles need to be built. For a given list and underlying principles, it needs to be clear whether a list is solely indicative of the underlying principle or is exhaustive.

3. **Regularly update lists based on new evidence.** A transparent process for regular, evidence-based updating of lists is required to address novel risks and changing circumstances. There is a need to develop processes to remove items from these lists; for example, when new vaccines or therapeutics become available or when a potentially pandemic pathogen becomes a pandemic pathogen.

4. **Determine triggers for updating.** More work could be done to identify biological components and/or combinations of components that create risk. These, in turn, could be used to identify and update lists for novel threats, including traditional lists of pathogens. For example, a list of potential pandemic pathogens could be automatically updated if an agent was found to exhibit a given human receptor binding affinity and lack of preexisting population immunity.

5. **Facilitate legitimate science.** Biosecurity tools that better match the potential to cause harm, and that are tailored for use with parts rather than whole organisms, tailored specifically for the purposes to which they are put, and updated continuously to keep pace with scientific understanding should reduce the restrictions or oversight burdens on scientists and researchers, compared with current approaches. Improving the fit between biosecurity tools and the concerns they are meant to address should facilitate legitimate science and not further restrict it.

Adapt Taxonomic Lists to Consider Ease of Engineering

The misuse potential of pathogens depends not only on their inherent pathogenicity but also on whether they might be engineered to do additional harm. As a result, it may be desirable to use taxonomy-based lists to identify not just pathogens, but also which pathogens may be more readily misused because they are easier to work with or to engineer.[37] For example, additional biosecurity oversight may be warranted for viruses with smaller genomes or those that are easier to "boot up" from naked genetic material to infectious agent. Our understanding of the science that underpins this is evolving rapidly. It will be vital that any oversight measures that consider the ease of engineering will need to be highly adaptive—they will need continual review and updating, as well as clear definitions and thresholds for what is meant by "ease" of engineering and how that ease is linked to increased risk of harm.

Clarify Language Used in Taxonomic Lists

Taxonomic lists that define entries too broadly may unintentionally capture organisms that pose little potential for harm. For instance, a hypothetical list entry for "influenza A viruses" would group together both viruses with human pandemic potential and strains with mild transmissibility. This capturing of unintended strains may inadvertently hinder public health research. For instance, Australia Group export controls and national export controls outside the United States extend to "SARS-related coronaviruses," potentially restricting researcher access to SARS-CoV-2 at a time when the virus was widely circulating in human populations and when countermeasure development was sorely needed. Clarifying the language used in taxonomic lists may assist implementers of biosecurity controls. For example, clarification as to whether SARS-CoV-2 is a "SARS-related coronavirus" for the purpose of export control has been made in some settings, such as by the US Department of Commerce, but is missing in other settings, such as the Australia Group common control lists.[38]

To avoid capturing large numbers of pathogens not relevant to biosecurity controls, some taxonomic lists have already had to develop additional criteria to focus controls beyond taxonomy. For example, attenuated strains used in many traditional vaccines have substantially different properties than their ancestors but still fall into the same taxonomic group. This has required clarification of rules. For example, the US Federal Select Agent Program features an avenue to apply for exclusions of attenuated strains of select agents, but these exclusions are granted on a case-by-case basis without clear guidelines that would enable automatic exclusion rather than review by the regulator.

Taxonomic lists that define entries too narrowly can fail to capture organisms that pose a harm equivalent to those listed. For example, the US Select Agent and Toxins list does not extend to all influenza A viruses but only captures 2 of its members ("1918 reconstructed influenza" and "highly pathogenic avian influenza viruses").[‡][39] These rules

---

[‡]Avian influenza strains are included for their danger to livestock, in particular chickens.

have failed to capture other related viruses that may pose similar harms. For example, research to make H2N2 (a pathogen that has already caused a pandemic) more transmissible or virulent would not require enhanced US federal oversight.[40]

Specify Controlled Taxons via Reference to Database Identifiers

Current lists often use species names (eg, "*Bacillus anthracis*" or "SARS-related coronavirus"), but databases attempting to capture the relationships between various taxonomic levels exist. They could serve as more specific reference points for these control lists. For example, entries in taxonomy databases, such as those maintained by the US National Center for Biotechnology Information, are hierarchical and very specific (eg, SARS-CoV-2 is taxonomy ID 2697049, which falls under taxonomy 694009 for SARS-related coronaviruses). If the US Department of Commerce really meant to control all SARS-related coronaviruses, expressing this as control of taxon 694009 provides greater clarity. Alternatively, if they only meant to control strains from the original SARS-CoV-1 and Middle East respiratory syndrome viruses, use of taxon IDs 2901879 and 1335626 would clearly indicate this intent. Furthermore, there have been efforts to develop taxonomic groups defined by reciprocal nucleotide identity, which streamlines and standardizes the definition of novel taxa and potentially increases their utility as part of a biosecurity control regime.[41,42]

MOVING FROM TAXONOMIC LISTS TO SEQUENCE OF CONCERN DATABASES

There are common constructs used in modern biotechnology that are not easy to classify using traditional taxonomy. Chimeric viruses, made up of fragments of different taxonomic origin, are frequently created for gene therapy, vaccines, and oncolytic viruses. Variations of such chimeric viruses could have pathogenic potential, yet multifamily chimeras may be excluded from the US Federal Select Agent Program even if they use select agent genes.[43] Indeed, associated guidance states that "chimeras that are comprised of select agent and nonselect agent genes from the same virus family require careful review to determine select agent status," admitting that chimeras defy taxonomic classification.[44] Assessing the potential harm posed by a chimeric or recombinant virus can prove challenging: 1 case report details an HIV-1 infection acquired during routine experiments with supposedly noninfectious pseudoviruses.[45] The report speculates that the worker was infected by an accidentally generated hybrid virion with expanded tropism and infectivity.

It may soon be possible to create novel organisms with potential for harm from "harmless" organisms without intentionally adding known genes of concern, whether through directed evolution[46] or new multiplexed genome editing technology.[47] Furthermore, some emerging technologies, such as *de novo* protein engineering, do not fit with an organism-based threat model and challenge taxonomic classification.

As it becomes increasingly possible to engineer organisms at the genetic level, sequence-based approaches may prove more effective than those based on whole organisms.

Improving Sequence Databases Used to Implement Taxonomic Lists

Although the carefully curated taxon IDs managed by the US National Center for Biotechnology Information provide a reliable reference set upon which to make taxonomic classifications, the companion sequence databases, such as GenBank, are neither curated to the same degree nor were they designed to enable classification of a sequence's taxonomic origin. They were created to be and continue to serve as a tool for researchers to exchange data.[27] As a result, these databases often include entries that have sequences that do not make taxonomic sense. For example, a researcher may upload a record consisting of a limited target sequence from a pathogen combined with the sequence of an unrelated reporter. The researcher may choose to label this record with the taxonomic identifier—for example, Ebola—perhaps because it is part of a set of similar sequences, each containing a snippet from a different pathogen. This means that other uses of this benign reporter gene may have a taxonomic best match to a database record classified as Ebola. Databases that provide reference genomes (ideally multiple references per organism) along with high-confidence individual gene records, explicitly for the purposes of sequence screening, would improve the accuracy and efficiency of taxonomic sequence screening while reducing the costs associated with addressing false positives.

Sequences of Known Toxins and Virulence Factors

Given the desire to determine potential for harm based on functional characteristics rather than taxonomic categorization, databases should be developed to evaluate potential for harm based on functions inferred from DNA and protein sequences.[1,3] Such a database would be immediately useful for screening synthetic DNA orders.[8,37] A database of sequences of concern for this context would include sequences that encode listed toxins and genes known to "endow or enhance" pathogenicity from regulated pathogens. This would help clarify whether a specific sequence is subject to control or not. For example, at present, nearly 70% of virulence factors found in pathogenic bacteria—those that "endow or enhance pathogenicity"—are also found in nonpathogenic organisms.[48] Such databases could be derived from existing, publicly available lists such as Victors[49] and the Virulence Factor Database.[50] The Nuclear Threat Initiative–World Economic Forum Technical Consortium for DNA Synthesis Screening is developing this type of database for its Common Mechanism.[32]

Capturing a Wider Range of Sequences of Concern

There has been considerable interest in developing more powerful databases of sequences of concern that can capture a much wider range of functions, including those that are not found in known toxins and pathogens. The revised *Screening Framework Guidance for Providers and Users of Synthetic Oligonucleotides*,[51] released in April 2022, recommends the use of such databases to flag potential harms from sequences of synthetic DNA, and several tools have been developed that could be used for this purpose.[31,52,53] There is also interest in further expanding these tools to include machine learning-enabled predictions that can identify potential harms in unknown sequences. Users could also choose to include DNA sequences related to harms beyond pathogenicity or toxicity, such as those that could enable antimateriel applications.

Limitations of Sequence of Concern Databases

Sequence of concern databases present their own challenges, especially related to curation and access. Sequence databases require expert curation both in their initial development and to keep pace with advances in pathogen research, which can be costly and require subjective judgments. Curation also requires clear annotations and communication about what sequences or functions should be considered of concern, by whom, and for what purpose (ie, regulatory control or merely increased attention on the ordering customer). An example of how this could be achieved is the Comprehensive Antibiotic Resistance Database, which combines manual expert curation (aided by machine learning-assisted text mining) with standardized vocabulary and inclusion criteria, such as requiring experimental validation of stated function, submitting the sequence to a public database, and publication on the mechanism in question.[54] Concerns about misuse of these tools and information hazards contained in these databases are especially relevant as they expand to capture broader harms than those currently encompassed by existing public-facing lists.

Enumerating Concerns in a Sequence- or Taxonomy-Agnostic Way

Biological systems can have emergent properties that pose concerns and are not easily described at the level of individual sequences. Some concerns arise from combinations of sequences. For example, sporulation is associated with weaponization potential because spores can be aerosolized and are resilient to harsh environmental conditions. Yet sporulation is a multistep biological process that relies on specific combinations of genes, making it difficult to capture through sequence-based approaches alone. In addition,

synthetic biology tools can be used to combine otherwise innocuous sequences into biological systems that may pose biosecurity concerns, such as gene drives. Oversight mechanisms that regulate research at the level of biological functions, whether present naturally or endowed experimentally, could help to address some scenarios where sequence- and taxonomy-based approaches fall short.[55] If research is regulated at the level of biological function, then lists, other guidelines, or algorithmic functional prediction tools may help review committees put these regulations in action.

A practical example of function-based approaches has been used in recent years for the International Genetically Engineered Machine (iGEM) competition. The competition uses a so-called "White List" to help identify biological materials that may pose a greater hazard and whose use requires prior approval and enhanced levels of oversight. In previous iterations, this approach addressed the use of specific organisms or biological parts in the competition. Since 2019, a series of activities has been included in the list requiring prior approval. Activities requiring additional hazard identification and risk management include all 7 experiments of concern, anything that notably alters risks associated with antimicrobial resistance, and anything that is likely to bias the inheritance frequency of a genetic marker in an organism's progeny, such as through the creation of a gene drive.[56] In all these cases, it is the likely outcome that is subject to a rule and not a specific organism, part, or sequence. These functional definitions have been reviewed and revised considering practical experience in their implementation.[57]

Machine learning has enabled a dramatic shift in our understanding of proteins. Recently, Meta's ESMFold model has allowed the prediction of protein structure for over 600 million protein sequences.[11] This, in turn, has enabled much greater fidelity in being able to design novel proteins for specific biological functions. Both being able to predict function from structure and being able to design for a specific function were once considered to be grand challenges. Progress in addressing these challenges is happening more quickly than anticipated. The same tools and approaches are now being applied to more complex biological challenges, such as use of "hallucination" to broaden potential protein design space, design of very small beta-barrel domains, or functional prediction of thermostability.[58-60] It is possible that advances in characterizing biological systems will open the door to predicting and designing biological functions more broadly, thereby enabling biosecurity tools that look beyond just taxonomy or sequence. For example, generative protein modeling opens the door to much-needed screening tools based on protein structure or predicted binding targets.

Considerations beyond biological function should also be incorporated into biosecurity regimes where relevant. The specific public health, environmental, and economic

impacts of a biological agent may be regionally and temporally specific. It may be undesirable or impossible to regulate access to a pathogen in areas where it is endemic or during an active global health crisis, and agricultural pathogens affecting certain crops may be less concerning in areas where that crop is not grown.

Any taxonomy- or sequence-agnostic approach should be carefully implemented to avoid introducing unmanageable levels of ambiguity. For example, whether research falls under the US potential pandemic pathogen care and oversight[18] federal guidance is based on function, determined by whether it can be "reasonably anticipated to involve creation, transfer, or use of enhanced [potentially pandemic pathogens]."[61] In this context, "reasonably anticipated" could mean either the researcher "wouldn't be surprised if" or alternatively "would be surprised if it did not." The former captures much more research than the latter. Guidance on which approach was intended would help reduce subjectivity and assist in ensuring relevant research is being considered. Efforts by the National Science Advisory Board for Biosecurity to revise guidance for US rules and regulations provide an important opportunity for ensuring greater clarity.[62]

Equally, taxonomy- or sequence-agnostic approaches will need to be carefully developed and implemented to address their own potential for misuse. In some cases, the tools developed to manage potential for harm could assist those set on causing harm. For example, a machine learning tool that predicts potential harm from a sequence might be repurposed to identify novel harms or paths that circumvent existing biosecurity controls. We must ensure that biosecurity tools keep pace with these capabilities as they become available.

Conclusion

We need to be more deliberative in the development and implementation of biosecurity regimes. Taxonomic lists of pathogens have provided a historical bulwark, but today, we have many more tools at our disposal. We must first understand the goal of the biosecurity regime and the biological functions that it is attempting to control; only then can we choose an appropriate tool to control these functions.

Biosecurity tools that better match the potential to cause harm, are tailored for use with parts rather than whole organisms and are updated continuously to keep pace with scientific understanding should reduce the oversight burdens on scientists compared with current approaches.

Taxonomic lists will remain important for certain purposes, such as controlling access to or criminalizing unauthorized possession of specific organisms. We have identified several ways in which taxonomic lists and their use inside biosecurity regimes can be improved.

We have also identified cases in which biosecurity regimes might make use of other types of lists. For example, efforts to control access to genetic elements or to manage risks associated with synthetic biology might benefit from databases of sequences of concern, both for implementing existing controls of known toxins and pathogens and for capturing broader functions of concern.

Ultimately, it may be desirable to move away from lists of organisms or sequences to more broadly describe the functions that are being controlled. There have already been efforts to integrate specific functions into biosecurity regimes. New tools are emerging to enable increasingly sophisticated implementation of function-based rules. Testing the efficacy of these tools, fostering agreement as to their utility, integrating them into regulations and oversight systems, and implementing their use will be challenging. We will need much more adaptive forms of risk management. There are notable regulatory and policy hurdles, as well as technical challenges, to be overcome.

In all cases, there is more research to be done. Taxonomic lists cannot be our only tool, especially as our understanding of biology and what disrupts its normal functioning grows. We can, and must, build better tools that are tailored to the realities of modern biotechnology.

References

1. Casadevall A, Relman DA. Microbial threat lists: obstacles in the quest for biosecurity? *Nat Rev Microbiol*. 2010;8(2):149-154.
2. Méthot P-O. Science and science policy: regulating "select agents" in the age of synthetic biology. *Perspect Sci*. 2015;23(3):280-309.
3. National Research Council. *Sequence-Based Classification of Select Agents: A Brighter Line*. Washington, DC: The National Academies Press; 2010. Accessed June 9, 2023. https://doi.org/10.17226/12970
4. Petro JB, Plasse TR, McNulty JA. Biotechnology: impact on biological warfare and biodefense. *Biosecur Bioterror*. 2003;1(3):161-168.
5. World Health Organization (WHO). *Global Guidance Framework for the Responsible Use of the Life Sciences: Mitigating Biorisks and Governing Dual-Use Research*. Geneva: WHO; 2022. Accessed June 9, 2023. https://www.who.int/publications/i/item/9789240056107
6. Intelligence Advanced Research Projects Activity. Fun GCAT: Functional Genomic and Computational Assessment of Threats. Accessed June 9, 2023. https://www.iarpa.gov/index.php/research-programs/fun-gcat
7. Intelligence Advanced Research Projects Activity. FELIX: Finding Engineering-Linked Indicators. Accessed June 9, 2023. https://www.iarpa.gov/research-programs/felix
8. Diggans J, Leproust E. Next steps for access to safe, secure DNA synthesis. *Front Bioeng Biotechnol*. 2019;7:86.
9. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589.
10. Watson JL, Juergens D, Bennett NR, et al. De novo design of protein structure and function with RFdiffusion. *Nature*. 2023;620(7976):1089-1100.

11. Lin Z, Akin H, Rao R, et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*. 2023;379(6637):1123-1130.

12. World Health Organization (WHO). *Laboratory Biosafety Manual*. 4th ed. Geneva: WHO; 2020. Accessed August 31, 2023. https://www.who.int/publications/i/item/97892400 11311

13. Kojima K, Booth CM, Summermatter K, et al. Risk-based reboot for global lab biosafety. *Science*. 2018;360(6386):260-262.

14. Government of Canada. Risk groups and risk assessment. Updated February 16, 2023. Accessed August 8, 2022. https://www.canada.ca/en/public-health/services/laboratory-biosafety-biosecurity/risk-groups-risk-assessment.html

15. BAuA: Federal Institute for Occupational Safety and Health. Classification of biological agents in risk groups. Accessed August 8, 2022. https://www.baua.de/EN/Topics/Work-design/Biological-agents/Classification.html

16. National Health Commission of the People's Republic of China (NHC). *General Guidelines for Biosafety in Micro-biological and Biomedical Laboratories*. Beijing: NHC; 2002. Accessed August 8, 2022. http://www.nhc.gov.cn/wjw/s9502/201212/34016.shtml

17. Public Health Emergency. Recommended policy guidance for departmental development of review mechanisms for potential pandemic pathogen care and oversight (P3CO). Published January 9, 2017. Accessed August 31, 2023. https://www.phe.gov/s3/dualuse/documents/p3co-finalguidancestatement.pdf

18. US National Institutes of Health. Research involving enhanced potential pandemic pathogens. Last reviewed June 5, 2023. Accessed June 12, 2023. https://www.nih.gov/news-events/research-involving-potential-pandemic-pathogens

19. Dance A. The shifting sands of 'gain-of-function' research. *Nature*. 2021;598(7882):554-557.

20. Australia Group. List of human and animal pathogens and toxins for export control. Published November 30, 2022. Accessed June 27, 2022. https://www.dfat.gov.au/publications/minisite/theaustraliagroupnet/site/en/human_animal_pathogens.html

21. Palmer MJ. Learning to deal with dual use. *Science*. 2020;367(6482):1057.

22. Inglesby TV, Lipsitch M. Proposed changes to U.S. policy on potential pandemic pathogen oversight and implementation. *mSphere*. 2020;5(1):e00990-19.

23. Permanent Mission of the People's Republic of China to the United Nations and Other International Organizations in Vienna. Dual-use biological agents and related equipment and technologies export control list. Published January 14, 2002. Accessed August 8, 2022. http://vienna.china-mission.gov.cn/eng/hplyhnyfks/200201/t20020114_8850602.htm

24. Effective Prosecution of Possession of Biological Toxins and Agents Act of 2019, Pub L No. 116-31, 133 Stat 1034-1035 (2019).

25. US Federal Select Agent Program. Select agents and toxins list. Last reviewed August 1, 2023. Accessed August 8, 2023. https://www.selectagents.gov/sat/list.htm

26. Mackelprang B, ed. Security screening in synthetic DNA synthesis: recommendations for updated federal guidance. Engineering Biology Research Consortium. Published April 2022. Accessed June 9, 2023. https://ebrc.org/wp-content/uploads/2022/04/EBRC-2022-Security-Screening-in-Synthetic-DNA-Synthesis.pdf

27. Beal J, Clore A, Manthey J. Studying pathogens degrades BLAST-based pathogen identification. *Sci Rep*. 2023;13:5390.

28. United States Government. United States government policy for oversight of life sciences dual use research of concern. Published March 2012. Accessed October 6, 2023. https://www.phe.gov/s3/dualuse/documents/us-policy-durc-032812.pdf

29. Ritterson R, Kingston L, Fleming AEJ, Lauer E, Dettmann RA, Casagrande R. A call for a National Agency for Biorisk Management. *Health Secur*. 2022;20(2):187-191.

30. SecureDNA. Cryptographic aspects of DNA screening. Accessed April 29, 2023. https://securedna.org/research/

31. Balaji A, Kille B, Kappell AD, et al. SeqScreen: accurate and sensitive functional screening of pathogenic sequences via ensemble learning. *Genome Biol*. 2022;23(1):133.

32. Nuclear Threat Initiative, World Economic Forum (WEF). *Biosecurity Innovation and Risk Reduction: A Global Framework for Accessible, Safe and Secure DNA Synthesis*. Geneva: WEF; 2019. Accessed July 25, 2022. https://www.nti.org/wp-content/uploads/2020/01/Biosecurity_Innovation_and_Risk_Reduction.pdf

33. Salter SJ. You cannot *B. cereus*. *Nat Rev Microbiol*. 2011;9(2):83.

34. Klee SR, Brzuszkiewicz EB, Nattermann H, et al. The genome of a *Bacillus* isolate causing anthrax in chimpanzees combines chromosomal properties of *B. cereus* with *B. anthracis* virulence plasmids. *PLoS One*. 2010;5(7):e10986.

35. Sandbrink JB, Alley EC, Watson MC, Koblentz GD, Esvelt KM. Insidious insights: implications of viral vector engineering for pathogen enhancement. *Gene Ther*. 2023;30(5):407-410.

36. Anti-terrorism, Crime and Security Act 2001, Schedule 5 Pathogens and Toxins (UK, 2001). Accessed August 8, 2022. https://www.legislation.gov.uk/ukpga/2001/24/schedule/5

37. Carter SR, Friedman RM. *DNA Synthesis and Biosecurity: Lessons Learned and Options for the Future*. La Jolla, CA: J. Craig Venter Institute; 2015. Accessed August 8, 2022. https://www.jcvi.org/research/dna-synthesis-and-biosecurity-lessons-learned-and-options-future

38. Millett P, Rutten P. COVID-19, SARS-CoV-2, and export controls. *Health Secur*. 2020;18(4):329-334.

39. US Centers for Disease Control and Prevention (CDC) and US Department of Agriculture Animal and Plant Health Inspection Service (APHIS). *2020 Annual Report of the Federal Select Agent Program*. Atlanta, GA and Washington, DC: CDC, APHIS. Accessed June 9, 2023. https://www.selectagents.gov/resources/publications/docs/FSAP_Annual_Report_2020_508.pdf

40. National Research Council. *A Survey of Attitudes and Actions on Dual Use Research in the Life Sciences: A Collaborative Effort of the National Research Council and the American Association for the Advancement of Science*. Washington, DC: The National Academies Press; 2009. Accessed June 9, 2023. https://doi.org/10.17226/12460

41. Tian L, Huang C, Mazloom R, Heath LS, Vinatzer BA. LINbase: a web server for genome-based identification of prokaryotes as members of crowdsourced taxa. *Nucleic Acids Res*. 2020;48(W1):W529-W537.

42. Vinatzer BA, Weisberg AJ, Monteil CL, Elmarakeby HA, Sheppard SK, Heath LS. A proposal for a genome similarity-

based taxonomy for plant-pathogenic bacteria that is sufficiently precise to reflect phylogeny, host range, and outbreak affiliation applied to *Pseudomonas syringae sensu lato* as a proof of concept. *Phytopathology*. 2017;107(1):18-28.

43. Possession, use, and transfer of select agents and toxins—addition of *Bacillus cereus* Biovar *anthracis* to the HHS list of select agents and toxins. *Fed Regist*. 2016;81(178):63138-63143. To be codified at 42 CFR §73.

44. US Federal Select Agent Program. Guidance on the regulation of select agent and toxin nucleic acids: regulated materials. Last reviewed January 18, 2023. Accessed September 13, 2023. https://www.selectagents.gov/compliance/guidance/nucleic/regulated.htm

45. Soria A, Alteri C, Scarlatti G, et al. Occupational HIV infection in a research laboratory with unknown mode of transmission: a case report. *Clin Infect Dis*. 2017;64(6):810-813.

46. Schmidt F, Weisblum Y, Rutkowska M, et al. High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape. *Nature*. 2021;600(7889):512-516.

47. Inscripta. Genome engineering and biosecurity: industry practices, resources, and opportunities: workshop report. Published February 2022. Accessed August 8, 2022. https://www.inscripta.com/wp-content/uploads/2023/03/Genome-Engineering-Biosecurity-Workshop-Report-1002771.pdf

48. Niu C, Yu D, Wang Y, et al. Common and pathogen-specific virulence factors are different in function and structure. *Virulence*. 2013;4(6):473-482.

49. Sayers S, Li L, Ong E, et al. Victors: a web-based knowledge base of virulence factors in human and animal pathogens. *Nucleic Acids Res*. 2019;47(D1):D693-D700.

50. Chen L, Yang J, Yu J, et al. VFDB: a reference database for bacterial virulence factors. *Nucleic Acids Res*. 2005;33(suppl 1):D325-D328.

51. Screening framework guidance for providers and users of synthetic oligonucleotides. *Fed Regist*. 2022;87(83):25495-25499.

52. Godbold GD, Kappell AD, LeSassier DS, Treangen TJ, Ternus KL. Categorizing sequences of concern by function to better assess mechanisms of microbial pathogenesis. *Infect Immun*. 2022;90(5):e0033421.

53. Adalja AA, Watson M, Toner ES, Cicero A, Inglesby TV. Characteristics of microbes most likely to cause pandemics and global catastrophes. In: Inglesby TV, Adalja AA, eds. *Global Catastrophic Biological Risks*. Cham, Switzerland: Springer Nature Switzerland; 2019:1-20. Accessed June 12, 2023. https://doi.org/10.1007/82_2019_176

54. Alcock BP, Huynh W, Chalil R, et al. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res*. 2023;51(D1):D690-D699.

55. National Research Council. *Biotechnology Research in an Age of Terrorism*. Washington, DC: The National Academies Press; 2004. Accessed June 12, 2023. https://doi.org/10.17226/10827

56. International Genetically Engineered Machine Foundation. White List. Accessed November 2, 2022. https://responsibility.igem.org/guidance/white-list

57. Millett P, Alexanian T, Palmer MJ, Evans SW, Kuiken T, Oye K. iGEM and gene drives: a case study for governance. *Health Secur*. 2022;20(1):26-34.

58. Anishchenko I, Pellock SJ, Chidyausiku TM, et al. De novo protein design by deep network hallucination. *Nature*. 2021;600(7889):547-552.

59. Kim DE, Jensen DR, Feldman D, et al. De novo design of small beta barrel proteins. *Proc Natl Acad Sci U S A*. 2023;120(11):e2207974120.

60. Frazer J, Notin P, Dias M, et al. Disease variant prediction with deep generative models of evolutionary data. *Nature*. 2021;599(7883):91-95.

61. US Department of Health and Human Services (HHS). *Framework for Guiding Funding Decisions about Proposed Research Involving Enhanced Potential Pandemic Pathogens*. Washington, DC: HHS; 2017. Accessed July 27, 2022. https://www.phe.gov/s3/dualuse/Documents/P3CO.pdf

62. Pannu J, Palmer MJ, Cicero A, Relman DA, Lipsitch M, Inglesby T. Strengthen oversight of risky research on pathogens. *Science*. 2022;378(6625):1170-1172.

Address correspondence to:
Piers Millett
International Biosecurity and Biosafety Initiative for Science
1776 I Street NW, Suite 600
Washington, DC 20006

Email: piers@ibbis.bio