NATIONAL ACADEMY OF SCIENCES
NATIONAL ACADEMY OF ENGINEERING
INSTITUTE OF MEDICINE
NATIONAL RESEARCH COUNCIL

REPORT IN BRIEF

Sequence-Based Classification of Select Agents: A Brighter Line

Although measures are in place to protect against the loss, theft, or misuse of Select Agents—a designated list of named pathogens and biological toxins with the potential to pose a severe threat to public health and safety or agriculture—concerns are growing that advances in molecular biology have created new biosecurity challenges. In particular, DNA synthesis technology may make it possible to construct Select Agents or modify them by introducing small changes to the genetic sequence. This report considers the scientific advances necessary to develop an oversight system for these "synthesized" Select Agents and other potential pathogens based on the predicted features and properties of the genes encoded by their DNA.

Recent years have brought rapid advances in biotechnology with clear benefits to humankind—but also the potential that these advances may be used for nefarious purposes. Of particular concern are the pathogens and biological toxins designated as Select Agents, identified as those posing the greatest potential threat to public health and safety or to domestic agriculture.

A photomicrograph of Bacillus anthracis bacteria using Gram-stain technique. Courtesy: CDC

Several universities, institutions, and government agencies maintain stocks of Select Agents for research purposes, and regulations are in place to prevent the loss, theft, or misuse of these materials; for example by restricting access to facilities that deal with Select Agents and ensuring adequate physical security measures. However, expanded access to DNA synthesis technologies—some 50 companies worldwide now offer the service at rapidly falling costs— is making it easier for researchers, industry scientists, and amateurs to construct organisms without needing to obtain samples of existing stocks or cultures. Concerns are growing that these DNA synthesis technologies could be used to synthesize Select Agents, to modify known Select Agents by introducing small changes to the genetic sequence, or to construct entirely novel pathogens.

In this context, the National Institutes of Health requested that the National Research Council convene a committee of experts to investigate the science and technology needed to replace the current Select Agent list with an oversight system that predicts if a DNA

What Are Select Agents?

Select Agents are a federally regulated list of bacteria, viruses, fungi, and toxins deemed to pose a significant threat to public health and safety, agriculture, the economy, or other factors. When a new agent emerges, it is named and research is initiated to study its biological makeup, the potential threat it presents, and its susceptibility to countermeasures— information that may be considered when deciding if an agent will be included on the Select Agents list. Currently, Select Agents are defined by their names, but it is not always clear which DNA sequences belong to a given name. Furthermore, Select Agents and non-Select Agents may have very similar DNA sequences.

sequence would result in an organism that should be regulated as a Select Agent. This approach would allow the rapid identification of Select Agents that are generated with DNA synthesis techniques, even though natural variation or intentional genetic modification blur the boundaries of a discrete Select Agents list based on a list of names.

After reviewing the current state of the science, the committee concluded that replacing the existing list of Select Agents with a system that could predict Select Agent status by DNA sequence alone is not feasible. However, the report says that DNA sequences could be used to establish a classification system to better define Select Agents in terms of DNA sequence, and to better monitor potential pathogens that could be produced using synthetic biology and synthetic genomics.

Table 1 Prospects for de novo prediction of "Select Agent-ness" from sequence

Property	Predictable now?	Foreseeable future?	Maybe someday?	Never
Pathogenicity			○	
Transmissibility			○	
Available treatments			○	
Ease of preparation			○	
Ease of dissemination			○	
Public perception				●
Historical bioweapon				●
Economic impact				●
Natural prevalence				●

Problems with Prediction

Select Agent status depends on both biological and non-biological characteristics. Non- biological factors—such as the potential economic impact of an agricultural pathogen or the potential for social disruption due to public perception-- are not part of an organism's genome sequence, and, therefore, it is not feasible to predict Select Agent status from a genome sequence alone.

Moreover, identifying the biological properties of a dangerous pathogen from a gene sequence is an extraordinarily difficult problem, requiring a deep understanding of the interaction of the complete pathogen and its host in the environment—a level of understanding that is currently impossible. For the foreseeable future, the only reliable predictor of the hazard posed by a biological agent is actual experience with that agent.

Fortunately, the threat of using synthetic biology to design entirely novel pathogens is also improbable—the designer would not be able to predict if the designed sequence would have pathogenic properties. In fact, the report's authoring committee cautioned against any security driven research effort dedicated to predicting if DNA sequence encodes a pathogen, because this type of information could empower future "bad actors" to use synthetic biology to develop new pathogens.

Gene Sequence-Based Classification System

It is likely that any threat from synthetic biology will come from the synthesis of known Select Agents or variants of them. To produce an oversight system equipped to respond to these challenges, the committee suggested modernizing the Select Agent regulations to define Select Agents in terms of their DNA sequences, rather than relying solely on a list of names of Select Agent organisms and toxins. This sequence-based classification system would be used to determine unambiguously if any genome sequence—for example, a sequence ordered from a DNA synthesis company—falls under the Select Agent Regulations, thus establishing a brighter line between Select Agents and non-Select Agents.

Important considerations for such a system include deciding how much and which sections of the gene sequence must be present to distinguish a Select Agent from a similar, but harmless, DNA sequence not covered by the Select Agent Regulations. For example, large parts of the genome of most organisms consist of DNA that is not related

Brighter line: Sequence-based classification does not add new names to the Select Agent list. The classification system would better define, in terms of DNA sequence, what is meant by each name on the existing Select Agent list.

Near-Term Scientific Milestones

The committee identified specific milestones that would aid in developing and implementing a sequence-based classification system, and could yield information to improve abilities to predict function of DNA from its sequence, and to improve understanding of infectious disease, including;

- *A sequence database with Select Agent focus*
  To establish a sequence-based classification system, it would be necessary to have a number of representative sequences that are classified as Select Agents, and a number of closely related sequences that are not. The database would also include naturally occurring genetic variation based on geographic distribution, ecological or laboratory adaptations.
- *An expanding sequence database of all biology*
  There are massive gaps in knowledge of the genetic characteristics of much of the biological world. A sequence database could help identify sequences of concern.
- *Define the criteria for Select Agent designation*
  The criteria designating a pathorgen or toxin as a Select Agent should be reviewed and clearly defined to allow unamibuous implementation of the Select Agent regulations. It would be difficult to create any clear and effective sequence based system, whether classification based or prediction based, if the criteria and purpose of the Select Agent list remain unclear.
- *Stratification of the Select Agent list*
  Several recent advisory panels have recommended the stratification of Select Agent lists to focus the highest scrutiny on those agents that are of the greatest concern. The report's authoring committee is in agreement with this, and concludes that it would make any sequence-based approach to oversight more feasible.

to pathogenicity—should these sequences from Select Agents, although presumed to have no role in disease, still fall under Select Agent regulations? To help answer this question, a sequence-based classification system would define a minimal parts list of the essential DNA sequences that make up the infectious form of each Select Agent.

Another consideration is determining how modifications to the DNA sequence of a Select Agent could affect its Select Agent status. Modifications may mean a DNA sequence is no longer recognized as that of a known pathogen, but it could still act like a Select Agent. How similar must the new sequence be in order to be considered a Select Agent? A successful classification system would include the most likely modifications that experts believe would likely also act as Select Agents, without including known non-Select Agents, such as vaccine strains. To remain current, the classification system would be updated frequently with new information to reflect the advances in biology, computation, and changes to the name-based Select Agent list.

A Yellow-Flag System

Although the classification system would help provide a DNA sequence-based definition of Select Agents—a brighter line for regulatory purposes—it would not address concerns about potentially dangerous sequences that are not themselves Select Agents. For example, an individual could order the genome of a Select Agent in pieces from different companies, or introduce just enough changes to evade Select Agent status. The report suggests that a yellow-flag biosafety system would work with sequence-based classification to flag sequences of concern. These might be genetic fragments of Select Agents or suspicious sequences that might encode potentially hazardous genes, yet are not themselves Select Agents. Upon identification of such a sequence, a yellow flag would be raised triggering a common-sense follow-up; for example, a DNA synthesis company may contact a customer to ensure that they know they have just ordered a sequence that could be considered dangerous, or pass the details of the order on to law enforcement authorities. The yellow flag biosafety system would evolve continuously and would not provide definitive information, making it inappropriate for regulatory implementation—but it could serve as a dynamic and timely resource for researchers,

Yellow-flag system: Sequence-based classification could also be used to help identify "sequences of concern" and serve as a resource for information sharing.

clinicians, amateur biologists, and for the law enforcement officials who monitor biological agents.

Recommendations

The sequence-based classification system is technologically feasible, and may improve the current system; however, such a system does have limitations and potential negative consequences and could be costly to establish. Therefore, the committee does not specifically recommend the implementation of a sequence-based classification system, but rather makes two recommendations:

- Boundaries should be defined to help determine whether a genome sequence is close enough to that of a Select Agent to fall under the Select Agent Regulations so that Select Agent status can be unambiguously determined from a genome sequence (for example, by a DNA synthesis company). These boundaries should be broad enough to include the plausible modifications that experts reasonably believe would probably also act as Select Agents, without encompassing existing non-Select Agents.

Researchers working at a large-scale DNA sequencing center.
Courtesy: The Broad Institute of MIT and Harvard

- A sequence-based classification system could address this problem. The advantages of the system should be considered and weighed against the costs and complexity of implementing the new system.

Committee on Scientific Milestones for the Development of a Gene-Sequence-Based Classification System for the Oversight of Select Agents: James W. LeDuc, (*Chair*), Galveston National Laboratory and The University of Texas Medical Branch, Galveston; **Ralph Baric,** University of North Carolina, Chapel Hill; **Roger G. Breeze,** Centaur Science Group; **R. Mark Buller,** Saint Louis University School of Medicine; **Sean R. Eddy,** Howard Hughes Medical Institute; **Stanley Falkow,** Stanford University School of Medicine; **Rachel Levinson,** Arizona State University; **John Mulligan,** Blue Heron Biotechnology; **Alison O'Brien,** Uniformed Services University of Health Sciences; **Francisco Ochoa-Corona,** Oklahoma State University; **Jane S. Richardson,** Duke University Medical Center; **Margaret Riley,** University of Massachusetts; **Tom Slezak,** Lawrence Livermore National Laboratory; **India Hook-Barnard** (*Study Director*), **Carl-Gustav Anderson,** *(Senior Program Assistant),* **National Research Council.**

The National Academies appointed the above committee of experts to address the specific task requested by the National Institutes of Health. The members volunteered their time for this activity; their report is peer-reviewed and the final product signed off by both the committee members and the National Academies. The members volunteered their time for this activity; their report is peer-reviewed and the final product signed off by both the committee members and the National Academies. This report brief was prepared by the National Research Council based on the committee's report.

For more information, contact the Board on Life Sciences at (202) 334-2187 or visit http://dels.nas.edu/bls. Copies of *Sequence-Based Classification of Select Agents: A Brighter Line* are available from the National Academies Press, 500 Fifth Street, NW, Washington, D.C. 20001; (800) 624-6242; www.nap.edu.