

# Natural or not? Scientists aid in quest to identify genetically engineered organisms

August 14 2023, by Aliyah Kovner

---



Credit: Pixabay/CC0 Public Domain

Ever since gene editing became feasible, researchers and health officials have sought tools that can quickly and reliably distinguish genetically modified organisms from those that are naturally occurring. Though

scientists can make these determinations after careful genetic analysis, the research and national security communities have shared a longstanding unmet need for a streamlined screening tool. Following the emergence of SARS-CoV-2, the world at large became aware of this need.

Now, such tools are being built.

A suite of techniques—one lab-based platform and four computational DNA sequence analysis models—was developed and refined over the course of a six-year program funded by the United States Intelligence Advanced Research Projects Activity (IARPA). These approaches have the potential to dramatically shift current screening capabilities for detecting engineered organisms.

Susan Celniker's team at Lawrence Berkeley National Laboratory (Berkeley Lab) was chosen to lead the testing and evaluation phase of the program, called Finding Engineering-Linked Indicators, or FELIX. She and her colleagues designed and produced increasingly challenging [biological samples](#) and assessed how well the tools made by participating academic and industry groups performed.

"What the FELIX program revealed in its initial months was that the capability to efficiently identify modified organisms in the environment does not exist. And so, the program really started at the foundations to developing first-in-class capabilities to identify modified organisms," said Ben Brown, a staff scientist computational biologist in Berkeley Lab's Biosciences Area, who co-led the project design with Celniker. "It's a very important program in that it created the tools to fill an important segment of our national security space."

## Testing the testers

To evaluate the work accomplished by its research teams, IARPA leveraged national laboratories to perform Test and Evaluation. This process ensures capabilities and tools that are developed under programs like FELIX can achieve the same results as reported by the researchers and are meeting program metrics, enabling evaluation of progress within the program. To ensure the tests would be as useful as possible for national security applications, the teams evaluated their performance with samples based on current and potential real-world scenarios.

"We got a list of every virus and microbe that people are worried about, and they went into the samples. The idea is that these testing systems will be prepared for a situation where it becomes necessary to confidently evaluate if an organism, be it mammalian, plant, microbe, or virus, has been engineered and is now circulating in the environment uncontained," says Celniker.

In total, the scientists at Berkeley Lab, Pacific Northwest National Laboratory, and the United States Department of Agriculture produced nearly 200 unique sample organisms with innocuous modifications ranging from large DNA sequence deletions or insertions all the way down to very subtle single nucleotide alterations made using CRISPR. Each testing group was given samples containing altered organisms as well as unmodified control samples containing non-modified organisms—known as "wild type"—that had never been fully sequenced before, so the genomes were not available in any database for comparison. The samples included virus particles and cells from bacteria, mammals, and fungi. These blinded samples represented potential human pathogens, such as HIV and E. coli, plant-infecting pathogens, and engineered complex species. To ensure health and security for participants, all of the microbial or viral samples created for testing were noninfectious and all were controlled under strict biosafety procedures.

The Testing and Evaluation portion of FELIX was divided into four phases, where each subsequent phase had more difficult samples. Groups with candidate tests were eliminated along the way if their technique did not perform well enough.

In the beginning, testing groups received purified samples with only one organism each, and they got multiples of every sample to determine whether the testing technique generated reproducible results. At the end, the testers received mixed samples designed to approximate real-world testing conditions. "For the final round, we gave them mixtures of up to 10 wild type and engineered organisms with different mutations in them to mimic what a soil sample might look like. And we actually did give them two soil samples as well as actual microbiome samples from a cow digestive tract and a mouse digestive tract," said Celniker. "So they got very complex samples that were really challenging."

Celniker and Brown further challenged the testing groups by designing samples that incorporated naturally occurring genetic oddities. For example, they presented samples containing bacteria that had acquired [new genes](#) by swapping plasmids—circular pieces of DNA that are separate from the cell's main genome—with other species of microbes. Gene acquisition from plasmids is very common in single-celled organisms, and it is through this mechanism that strains of bacteria can very quickly gain new traits such as antibiotic resistance.

They also threw in some hybridized influenza samples that could not have formed naturally (despite the virus's penchant for genetic cross-over) because the strains never circulated at the same time or on the same continents. Real-world gene scrambling events like these make it difficult to differentiate between natural and synthetic gene additions, but being able to do so is an essential capability of a modified organism detection tool.

To that end, the IARPA program leaders set an ambitious goal for the testing technologies of 99% specificity (no more than 1% of wild types misidentified as modified) and 90% sensitivity (no more than 10% of tests could misidentify a modified organism as wild type). The four techniques that passed through to the end of phase four testing and will be useful for identifying biological threats were a lab-based test from the company Draper and computational models from Raytheon, Ginkgo Bioworks, and Noblis. These techniques were shown to be excellent at identifying wild type organisms, and a Berkeley Lab-developed ensemble of the computational models achieved 99% specificity.

The sensitivity in identifying engineered organisms of individual models was between 55%-70%. But the ensemble was able to achieve approximately 72% sensitivity under cross validation, which occurred when it was tested on new sequence datasets. Overall performance of individual models and the ensemble demonstrated considerable improvement over existing state-of-the-art capabilities.

## **A new resource**

One reason why it's so hard to tell natural and engineered organisms apart is that scientists around the world use many different databases and programs to review and store genome sequence data. And on top of that, people use different names and terms to describe genes and predict their functions based on the sequences—a process called annotation. So, despite the fact that more and more species have had their genomes sequenced, the data isn't necessarily easy to use.

To remedy this issue, Celniker recruited her Biosciences Area colleague Chris Mungall, a computer staff scientist, to lead the development of an open-access software program and database. The result was Synbio Schema, which catalogs the annotated genomes of national security-relevant engineered and wild type organisms using standardized

language. Each sample that Celniker's team created for the testers was also added to the new database and annotated with the standardized language, providing an easy-to-use resource for future researchers.

"This is the first curated database and common language for engineered vs. non-engineered organisms, and they really had to build the airplane in flight because nothing like it existed previously, and the program would have been crippled without it," Brown said.

"The real problem arises when multiple research groups are trying to share and compare results," explained Mark Miller, a software developer in Mungall's group. "If there are any internal inconsistencies or other issues within a team's database, or if there are structural or nomenclature differences between the teams' databases, then nobody can tell whether one team's data agrees with the other teams." This forces scientists to tediously review annotations manually for accurate comparisons.

## **Growing the biodefense industry**

Building on the success of the FELIX program, the Berkeley Lab scientists plan to expand the database by adding new organisms that could be exploited as bioweapons, and call on other groups to add new sequences as well. Meanwhile, Brown is looking forward to using the neatly organized database to train machine learning models, which will lead to even better modified organism detection tools in the future.

Looking to next steps, the team hopes to use the knowledge and techniques gained from the FELIX program to develop better detection tools capable of ecosystem-scale monitoring to detect threats in the environment in real time—a capability that Brown describes as "NORAD for biology."

Provided by Lawrence Berkeley National Laboratory

Citation: Natural or not? Scientists aid in quest to identify genetically engineered organisms (2023, August 14) retrieved 3 May 2024 from <https://phys.org/news/2023-08-natural-scientists-aid-quest-genetically.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.