BIOSECURITY

The fight to keep dangerous DNA out of terrorists' hands

Machine learning could help firms avoid making dangerous organisms on demand.

BY SARA REARDON

Biologists the world over routinely pay companies to synthesize snippets of DNA for use in the laboratory or clinic. But intelligence experts and scientists alike have worried for years that bioterrorists could hijack such services to build dangerous viruses and toxins — perhaps by making small changes in a genetic sequence to evade security screening.

Now, the US government is backing efforts that use machine learning to detect whether a DNA sequence encodes part of a dangerous pathogen. Researchers designing such artificial-intelligence-based screening tools are beginning to make progress, and several groups presented early results on 31 January at the American Society for Microbiology (ASM) Biothreats meeting in Arlington, Virginia.

Their findings could lead to a better understanding of how pathogens harm the body, as well as new ways for scientists to link DNA sequences to specific biological functions.

"In the past, you'd take the pathogen, lock it up and put an army in front of it and you'd be fine," says Omar Tabbaa, director of computational biotechnology at Battelle, a technology-development company in Columbus, Ohio.

But Tabbaa says that the decreasing cost and difficulty of DNA engineering has changed the nature of biosecurity threats. Anyone who wants a specific piece of DNA can have the string of letters, called bases, synthesized for pennies per base. In 2006, as a test, reporters at *The Guardian* newspaper in the United Kingdom paid a DNA-synthesis company to make part of the smallpox virus, prompting calls for stricter screening measures.

In 2009, several of the largest DNA-synthesis firms formed a consortium to create standardized procedures for checking sequences submitted by their customers against databases of known pathogens. If the automated screening flags up a sequence, the company can check whether the customer is a legitimate researcher before synthesizing the DNA.

But these existing programs pick out only the parts of sequences that exactly match those of known pathogens. A smart terrorist could fool the system by changing a few bases in DNA from a virus or a gene that produces a toxin, or even by designing an entirely new pathogen. Compounding the problem, the databases themselves are often riddled with errors.

ANNA SCHROLL/FOTOGLORIA/UIG VIA GETTY

Dangerous pathogens are kept in high-security labs.

With this in mind, in 2016, the US Intelligence Advanced Research Projects Agency (IARPA) launched an initiative to design better algorithms for spotting potentially threatening sequences. Five teams from industry and academia are competing in the programme, says its manager, John Julias. The agency declined to disclose the programme's budget.

DNA DERBY

By 2020, the teams are expected to have developed a way of determining, in less than two weeks, whether an unknown sequence poses a threat. That will be a difficult task, says Andrew Warren, a software engineer at the University of Virginia in Charlottesville. "We have to be able to recognize any organism on the planet and also its molecular function."

Warren's team is designing a program that compares 40 million records of sequences from 90,000 microbial species. The algorithm learns to recognize the DNA sequences of known toxins and pathogens, identifies their common characteristics and then searches for similar sequences in other organisms. It can already reliably predict which type of organism a sequence comes from, says Warren, whose team presented early results at the ASM meeting.

Tabbaa, whose team at Battelle is developing a similar algorithm using sequences from both public and proprietary databases, says that computer algorithms could recognize commonalities among pathogens that people would miss. That will help programs to distinguish between the important parts of a DNA sequence and those that can be changed without affecting the pathogen's function. The goal is to pinpoint sections that could pose a security threat in an unknown sequence.

The Battelle team hopes that the program could also reveal information about the basic biology of organisms — such as a universal DNA sequence that allows toxins or viruses to stick to cells. "We think there's a whole slew of things to come out of this," he says.

But Rob Carlson, managing director at Bioeconomy Capital, a venture-capital firm in Seattle, Washington, is sceptical that stopping DNA-synthesis companies from being exploited will prevent bioterror attacks. So far, most attacks have involved the release of lab-grown pathogens, he says; in 2001, for instance, 5 people in the United States died and 17 became ill after receiving anthrax-laced letters. He fears that any government efforts to regulate DNA synthesis would push would-be bioterrorists underground.

IARPA declined to comment on whether the agency shared such concerns. ■