# Could 'Smoking Gun' DNA Sequence Lead Detectives to Biological Terrorist?

*By Alex Berezow — August 8, 2018*



Credit: Storyblocks [1]

Just like fingerprints, we all have a unique set of behavioral quirks.

For example, I tend to drink triple shot, iced vanilla lattes. Before beginning my work, I clean off the table using water and a napkin. (Seriously, why are coffee shop tables always so disgusting?) And, oftentimes, I tip my glasses in a peculiar way as I write my articles.

None of these quirks is particularly unique. But taken together, I'm probably the only triple shot, iced vanilla latte-drinking, table-cleaning, glasses-tipping person in Seattle. If I ever committed a crime and the police were out to get me, this combination of quirks may be just enough to identify me.

The same is true of DNA. A lab that is studying, say the genetics of the bacterium *E. coli*, may engineer a strain by adding gene X and removing gene Y. By themselves, those genetic manipulations may be commonly performed in laboratories across the world, but in combination, they may not be. In such a way, "quirks" can be inadvertently introduced in the DNA that could help identify the bacterium as originating from one particular laboratory.

**Could a "Smoking Gun" DNA Sequence Lead Detectives to a Biological Terrorist?**

The advent of CRISPR and other genetic engineering techniques raises the possibility that rogue actors may be tempted to develop biological weapons, many of whom may wish to remain anonymous after deploying them. However, new research published in *Nature Communications* shows that, using deep learning algorithms, the lab-of-origin can be narrowed down substantially.

The authors, Alec Nielsen and Christopher Voigt of MIT, utilized convolutional neural networks --

typically used to classify images into various categories -- to identify telltale marks in DNA sequences. They fed the neural network "training data" that consisted of DNA sequences obtained from well-known laboratories.

Once the model was built, it had to be validated. This was done by feeding it yet more DNA sequences, but not ones that were used to build the model. In this way, the authors were able to determine if their neural network could consistently match DNA sequences with their labs-of-origin.

They were modestly successful. The neural network was able to correctly identify the lab-of-origin 48% of the time. (Though that may not sound impressive, guessing at random would yield a success rate of 0.12%.) Furthermore, 70% of the time, the neural network listed the correct lab in its top 10 predictions.

So the answer to the question posed in the headline is, "Not quite." But the authors have shown that deep learning methods may be important for national biodefense countermeasures.

Source [2]: Alec A. K. Nielsen & Christopher A. Voigt. "Deep learning to predict the lab-of-origin of engineered DNA." Nature Communications 9: 3135. Published online: 7-Aug-2018. DOI: 10.1038/s41467-018-05378-z

---

---