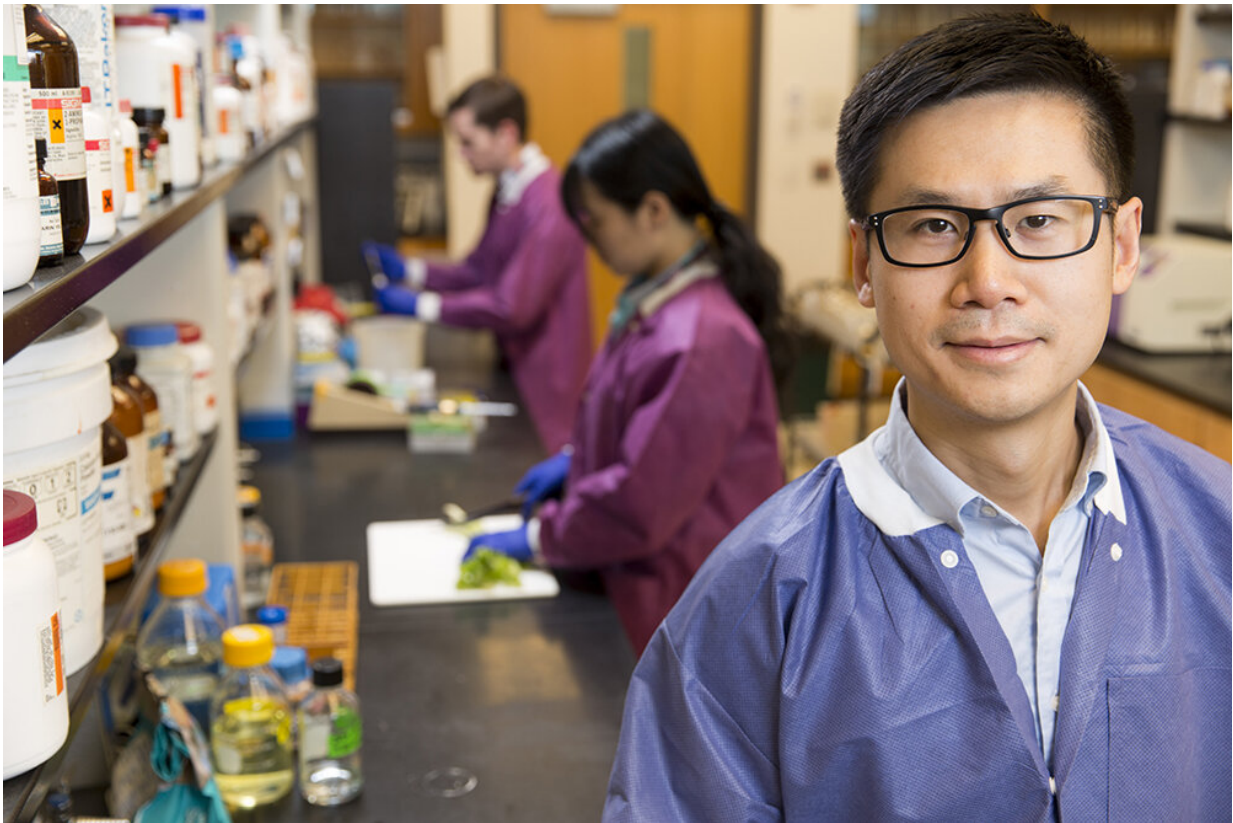


Scientists use machine learning to identify source of Salmonella outbreaks

February 11 2019, by Sharon Dowdy



UGA researchers Xiangyu Deng (shown) and Shaokang Zhang led a team of scientists who have trained an algorithm called Random Forest to predict certain animal sources of *S. Typhimurium* genomes. Credit: UGA

A team of scientists led by researchers at the University of Georgia

Center for Food Safety in Griffin has developed a machine-learning approach that could lead to quicker identification of the animal source of certain Salmonella outbreaks.

In the research, published in the January 2019 issue of *Emerging Infectious Diseases*, Xiangyu Deng and his colleagues used more than a thousand genomes to predict the animal sources, especially livestock, of *Salmonella Typhimurium*.

Deng, an assistant professor of food microbiology at the center, and Shaokang Zhang, a postdoctoral associate with the center, led the project, which also included experts from the Centers for Disease Control and Prevention, the U.S. Food and Drug Administration, the Minnesota Department of Health and the Translational Genomics Research Institute.

According to the Foodborne Disease Outbreak Surveillance System, close to 3,000 outbreaks of foodborne illness were reported in the U.S. from 2009 to 2015. Of those, 900—or 30 percent—were caused by different serotypes of Salmonella, including Typhimurium, Deng said.

"We had at least three outbreaks of Typhimurium, or its close variant, in 2018. These outbreaks were linked to chicken, chicken salad and dried coconut," he said. "There are more than 2,600 serotypes of Salmonella, and Typhimurium is just one of them, but since the 1960s, about a quarter of Salmonella isolates linked to outbreaks reported to U.S. national surveillance are Typhimurium."

The researchers trained the "machine," an algorithm called Random Forest, with more than 1,300 *S. Typhimurium* genomes with known sources. After the training, the "machine" learned how to predict certain animal sources of *S. Typhimurium* genomes.

For this study, the scientists used *Salmonella Typhimurium* genomes from three major surveillance and monitoring programs: the CDC's PulseNet network; the FDA's GenomeTrakr database of sources in the United States, Europe, South America, Asia and Africa; and retail meat isolates from the FDA arm of the National Antimicrobial Resistance Monitoring System.

"With so many genomes, machine learning is a natural choice to deal with all these data.

We used this big collection of Typhimurium genomes as the training set to build the classifier," said Deng who was awarded the UGA Creative Research Medal in 2017 for his work in this area. "The classifier predicts the source of the Typhimurium isolate by interrogating thousands of genetic features of its [genome](#)."

Overall, the system predicted the animal source of the S. Typhimurium with 83 percent accuracy. The classifier performed best in predicting poultry and swine sources, followed by bovine and wild bird sources. The machine also detects whether its prediction is precise or imprecise. When the prediction was precise, the machine was accurate about 92 percent of the time, Deng said.

"We retrospectively analyzed eight of the major zoonotic outbreaks that occurred in the U.S. from 1998 to 2013," he said. "The classifier attributed seven of them to the correct livestock source."

Deng says the tool has limitations; it cannot predict seafood as a source and it has difficulty predicting *Salmonella* strains that "jump around among different animals."

"I'd call this approach a proof of concept. It will get better as more genomes from various sources become available," he said.

In tweets about the study, Frank Yiannas, deputy director of the FDA, called the [machine learning](#) of whole genome sequences project "a new era of smarter [food safety](#) and epidemiology."

To the [average person](#), the success of this project means strains of *Salmonella Typhimurium* could be traced back to the source faster. Identifying what causes a foodborne illness outbreak is key to stopping it and preventing further illnesses.

"Using our method, investigators can better link cases of the same outbreak and better match isolates from food or food processing environments to isolates from sick people," he said. "This will give investigators more confidence to implicate a specific source that is behind the [outbreak](#)."

More information: Shaokang Zhang et al, Zoonotic Source Attribution of Salmonella enterica Serotype Typhimurium Using Genomic Surveillance Data, United States, *Emerging Infectious Diseases* (2018). [DOI: 10.3201/eid2501.180835](https://doi.org/10.3201/eid2501.180835)

Provided by University of Georgia

Citation: Scientists use machine learning to identify source of Salmonella outbreaks (2019, February 11) retrieved 24 April 2024 from <https://medicalxpress.com/news/2019-02-scientists-machine-id-source-salmonella.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.
