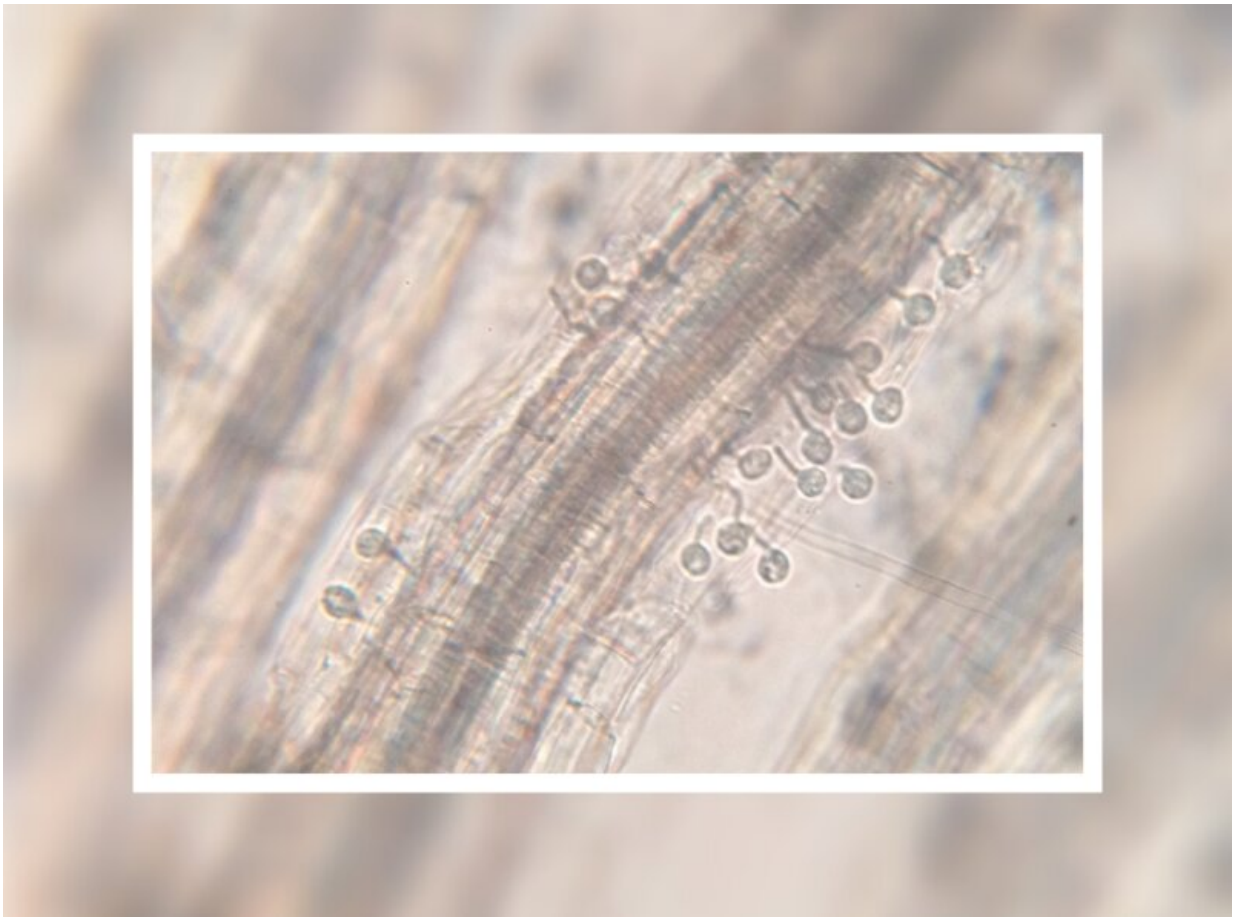# Researchers combine technologies to resolve plant pathogen genomes

January 29 2020, by Kendall Daniels

An image of P. capsici spores that are attached to an Arabidopsis root and initiating the process of penetration. Credit: John Herlihy for Virginia Tech.

With the help of new genomic sequencing and assembly tools, plant

scientists can learn more about the function and evolution of highly destructive plant pathogens that refuse to be tamed by fungicides, antibacterial, and antivirals.

But using these genomic technologies is not an easy task. The process not only requires time, but also money. In a recent paper published in *Molecular Plant-Microbe Interactions*, David Haak and John McDowell, from the School of Plant and Environmental Sciences in the College of Agriculture and Life Sciences, proved that these costly processes can be improved by combining two generations of technology.

What used to take a year-and-a-half and $2 million to complete can now be done within nine days for $1,000—and the technology performs with greater accuracy and field applicability than ever before.

"Think of it as analogous to a library full of books that are two-thirds or three-quarters completely written. What David has developed is a technology through which he could go to the library and finish those books really quickly and really accurately for a really low price point," said McDowell, the J. B. Stroobants Professor of Biotechnology.

Before this project began, Haak, an assistant professor, and his team had been trying to prove that it was possible to generate a completed assembly in a relatively short period of time—but they needed a relatively complex genome to test their theory. A few hallway conversations later, Haak and his students joined forces with McDowell and his team to unravel the complex genome of *Phytophthora capsici*.

"*P. capsici* is a representative of a really destructive group of pathogens. Its evolutionary cousin is the pathogen that caused the Irish Potato Famine in the 19th century, which killed at least a million people and caused at least a million more to relocate. These pathogens are still causing difficulty today," said McDowell. "One of the reasons for that is

because their genomes are exquisitely configured to enable them to evolve ways around interventions that farmers put in place in the field."

A microscopic image of a large root, which divides the photo into two diagonal sections. On each side of the root, there are many *P. capsici* spores, which resemble lollipops.

An image of *P. capsici* spores that are attached to an Arabidopsis root and initiating the process of penetration. John Herlihy for Virginia Tech.

In this species of pathogen, virulence genes are often located in gene-poor regions interspersed with repetitive regions within the genome. These repetitive regions are prone to rapid evolution and are the key to understanding its pathogenicity, or its ability to cause disease.

To better understand the inner workings of *P. capsici*, scientists must extract a DNA sample from the pathogen and perform genetic sequencing. Genetic sequencing is a process that determines the order of the nitrogenous bases—or the As, Cs, Gs, and Ts—that make up an organism's DNA.

However, genomic sequencing can read only a certain amount of DNA segments at one time. Scientists must then take these small sequences and re-assemble them so that the DNA is presented in the right order.

"Generating the sequence data, isn't really the problem. It's assembling that data. It's putting together the sequence information in the right order. The repeat-rich regions make us sometimes put two genes together that don't belong together or separate a full gene into two halves because we think a repeat goes right in the middle," said Haak.

All in all, resolving the genome of an organism requires powerful technology—and patience. And although bioinformatic technology has

made great leaps and bounds over the years, each generation isn't necessarily better than the last. Each generation of technology has its own forte.

Using first-generation technology, it would take one-and-a-half years and around $2 million to sequence the *P. capsici* genome. But with Haak's technology, it will take just nine days from DNA extraction to a polished assembly—and only cost $1,000. To make things even better, this technology will be able to sequence 100,000 times more information in roughly 1.5 percent of the time. And the technology is the size of a thumb drive.

Second-generation technology performs short read assemblies, which are extremely accurate; however, they do not span across repetitive regions well. And when scientists must go back and reassemble the genome, there is a reasonable chance of error.

"What happens with the short reads is that we don't know where those repeats begin and end, so we don't know where to put them to arrange them appropriately," said Haak.

Oxford Nanopore Technologies (ONT) MinION, or long-read sequencing, is the third generation of sequencing technology, but it has the opposite problem: it is far less accurate but it can give them a better overall picture by spanning across these critical repetitive regions.

Haak and his team combined these second- and third-generation technologies to exploit the accuracy of the former with the ability to span the repeated regions of the latter. It's the best of both worlds.

Upon using this new technology on *P. capsici*, Haak and McDowell got quite a shock. Haak and his group revealed that the genome is 1.5 times bigger than previously thought.

"That's 30 percent of the genome that we didn't even know existed, and that particular fraction of the genome is, undoubtedly, enriched with the sorts of genes that really make a difference in helping us understand what interacts with the plant or responds to fungicides or farmers' spray," said McDowell.

For Haak, the most exciting thing about the results of this paper is its proof-of-concept.

"We have something called the sequence archive database, which is full of all sorts of short-read sequences. We can actually leverage all of that existing data with this newer technology to be able to produce more genomes of this quality," said Haak.

Haak's new generation of technology is expected to revolutionize the way in which scientists collect genomic data. With their newly acquired, affordable, real-time data, scientists will be able to improve previous assemblies and quickly generate new ones that they can share to the sequence archives database. On a grander scale, this technology will advance the field of plant genomics and the worldwide effort to save the crop industry from destructive pathogens.

Now that Haak and McDowell have an estimated 97 percent of the genome for *P. capsici* in their grasps, they plan to use this information as supporting data for two new grant proposals. One proposal will focus on tomato and soybean diseases caused by pathogens of the Phytophthora group and the other proposal will focus on lavender, yet another victim of Phytophthora.

For Haak, this project was special because it was supported by a grant from the Fralin Life Sciences Institute at Virginia Tech with funds allocated to support the Global Systems Science Destination Area.

McDowell added, "I think it speaks to the environment here at Virginia Tech, promoted by Fralin, that enables these sorts of collaborations to come together and get some critical support in the early phase."

More information: Chenming Cui et al, Draft Assembly of Phytophthora capsici from Long-Read Sequencing Uncovers Complexity, *Molecular Plant-Microbe Interactions* (2019). [DOI: 10.1094/MPMI-04-19-0103-TA](DOI)

Provided by Virginia Tech