Next-Generation Sequencing Saves the Sweet Potato

Researchers at the ARC of South Africa use the MiSeq® System to identify and map sweet potato viruses, with the goal of enhancing food security in developing countries.

Introduction

Sweet potatoes are highly nutritious—a good source of carbohydrates, protein, fiber, iron, and vitamins A, B, and C—and are often used as poverty alleviation crops for food security in developing countries. Yet, viral diseases can reduce crop quality and yield by up to 100%.

With researchers at the Agricultural Research Council (ARC)[1] of South Africa, PhD student, Thulile Nhlapo, has studied DNA and RNA viruses affecting the sweet potato for the past 2 years. Using the MiSeq System, she has discovered and partially mapped 2 viruses never before identified in South Africa, and conducted metagenomic studies to fully sequence 5 other viruses. By designing primers to screen plants for these viruses, Ms. Nhlapo is enabling farmers to purchase and plant virus-free sweet potatoes in their fields.

iCommunity spoke with Ms. Nhlapo to learn about how she's identifying and mapping new sweet potato viruses using next-generation sequencing (NGS), and the impact this research will have on the agricultural community and food security programs around the world.

Q: Why is the sweet potato important as a food staple in the developing world?
Thulile Nhlapo (TN): The new orange-flesh sweet potato provides high-quality vitamin A, which is especially important for malnourished women and children. In fact, research has shown that the sweet potato aids in development and vision quality in children ages 3-7 years old. Many countries that depend on cassava as a primary food source will also use sweet potato as a supplement to enhance the nutritional quality of high-carbohydrate diets.

Q: What sparked your interest in agrigenomics research?
TN: I studied microbiology in college, where I had a broad spectrum of plants and animals to research. I'm not interested in animal work, because I don't like blood and moving things. So, I decided to pursue plant genomics instead.

Before starting this project, I was involved in some of ARC's outreach programs in the South African provinces that were focused on the benefits that sweet potatoes provide in the diets of vitamin A-deficient women and children. As a graduate student, I was handed a list of all the crops and animals I could study, and my familiarity with the sweet potato drew me to those projects.

My research is now focused solely on studying sweet potato viruses. I have the advantage of working alongside researchers who are looking at similar issues with the cassava in Mozambique and the Democratic Republic of the Congo, and the potato in South Africa.

Q: How is your work affiliated with the African Centre For Gene Technologies (ACGT)[2]?
TN: The ACGT is a consortium made up of the ARC, the Council for Scientific and Industrial Research (CSIR), the University of Johannesburg, the University of Pretoria, and the University of Witwatersrand. Its goal is to create a collaborative network of excellence in advanced biotechnology. I attend ACGT meetings with graduate students and established researchers once or twice a year, and presented the results of my sweet potato research at the 2014 meeting.[3] It is a good place to network and catch up on everyone else's research.

> With NGS technology, we can identify viruses in mixed infections and are able to differentiate viral strains.

Q: What technologies do you use to study sweet potato viruses?
TN: We use the MiSeq System. This technology enables me to sequence DNA and RNA samples directly and then identify known and novel virus genomes from the sequence assemblies. I have used this system to detect and study 7 sweet potato viruses in South Africa. There are about 30 sweet potato viruses that are known to exist worldwide.

Thulile Nhlapo is a PhD student at the Agricultural Research Council in South Africa.

Q: What are the advantages of NGS systems?
TN: The advantage of NGS is its efficiency and ability to look at a broader window of genomic information easily. For example, we were able to detect 5 viruses in a mixed infection in a single plant and detect 2 badnaviruses that had never been reported in South Africa before. A conventional polymerase chain reaction (PCR) or enzyme-linked immunosorbent assay (ELISA) assay would not detect those viruses, because you would need prior knowledge of their existence and the primer sequences and antibodies to aid in their detection. With NGS technology, we can identify viruses in mixed infections and are able to differentiate viral strains.

In one instance, we found that normal PCR could not identify a virus because the sequence was highly divergent. NGS enabled us to assemble the full genome of the virus and design primers from the sequence that we could use to screen the plants sold to farmers around the country.

Q: How are you using NGS to study these viruses?
TN: We performed a metagenomic study using the full microbial diversity in a sample. With metagenomics, you can look at bacterial, fungal, or viral contamination. I use viral metagenomics and perform whole-genome deep sequencing from total RNA after Ribo-Zero treatment, to assemble the full-length sweet potato RNA virus genomes. For DNA viruses, I use a rolling circle amplification (RCA) approach to enrich for the viral genomes. In the case of the 2 newly discovered viruses—Sweet potato badnavirus A and Sweet potato badnavirus B—we only managed to sequence partial genomes, but we will get additional sequence data to close the gaps.

> The biggest surprise in my research was detecting the 2 badnaviruses in South African sweet potato, because they had previously only been reported in Peru and Tanzania.

Q: How have you found the data quality of the MiSeq System?
TN: The MiSeq data quality is really good. To detect DNA viruses, we sequenced 27 RCA samples and obtained about 30 Mb of data per sample. For the detection of RNA viruses, we had to sequence slightly deeper—about 120–500 Mb of data per sample—which resulted in approximately 13 million reads in total for the RNA metagenome. This enabled us to assemble the viruses at very high sequence depths, about 500–700× coverage.

The biggest surprise in my research was detecting the 2 badnaviruses in South African sweet potato, because they had previously only been reported in Peru and Tanzania. We first detected a few reads mapping to the viruses in the RNA data. We then did small RNA sequencing and found over 50,000 siRNAs per virus. The power of NGS sequencing is unbelievable because it covers everything. If a virus is dominant in a plant, it will pick it up. Yet, it also picks up viruses that are present in the plant at low titers. That's very exciting.

After my presentation at the ACGT meeting last year, other researchers became interested in coming to our facility to perform NGS with the MiSeq System. They realized that they could get high coverage at a low cost-per-sample. The MiSeq System is beneficial for people performing viral research because even though the genomes are small, it delivers lots of data at high coverage depths, enabling us to assemble the full-length genome.

Q: What about the speed of your sequencing runs?
TN: I isolate the DNA or RNA, perform sample preparation, load the MiSeq System, and the turnaround time is quick. I do not have to wait long for data. Our sequencing runs take about 65 hours, and I'm pleased with that. I obtained a coverage depth of over 10,000× for one of the viruses and that was terrific. The sample was enriched for circular viral genomes by RCA prior to sequencing, hence this kind of coverage.

Q: What library prep kits are you using and what benefits do they deliver?
TN: For the RNA work, we use the TruSeq RNA Sample Preparation Kit and for the DNA work we use the Nextera Sample Preparation Kit. I also used the Ribo-Zero rRNA Removal kit (plant), which made it much easier to sequence viral genomic RNA within a total RNA sample.

Q: How are you performing data analysis?
TN: I used the CLC Bio Genomics Workbench version 7 for data analysis. I also used a tool for viral metagenomic analysis called MetaVir 2.

Q: What benefits will this research provide?
TN: The most immediate benefit is being able to design primers for multiplex PCR from our NGS-generated sequences. We can use these primers to screen plant material in our gene bank before we send it off to farmers and supply nurseries in all the 7 sweet potato growing provinces. This will enhance farmers' production yields, because these viruses have a significant negative effect on the yield and quality of a sweet potato crop.

> The MiSeq System is beneficial for people performing viral research because even though the genomes are small, it delivers lots of data at high coverage depths, enabling us to assemble the full-length genome.

Q: Do you have any scientific papers in development?
TN: I am finalizing the draft of a paper discussing the RNA viruses that I detected from deep sequencing with the MiSeq System and will be submitting it for publication soon. I am also working on a manuscript for my research with the DNA viruses.

Q: What are the next steps in your research?

TN: I am very excited about the 2 badnaviruses that we found. I plan to amplify the PCR products and sequence their full genomes. My next set of experiments will look at whether these 2 badnaviruses play any role in the infection of the plant, or whether they are latent and work in synergy with other viruses. I will also look at what phenotypic symptoms result when resistant and susceptible plants become infected with these viruses.

I would like to undertake expression analysis studies of sweet potato cultivars after infection with 2 RNA viruses (Sweet potato chloratic stunt virus and Sweet potato feathery mottle virus), 2 geminiviruses (Sweet potato mosaic associated virus and Sweet potato leaf curl virus) and the 2 badnaviruses (Sweet potato badnavirus A and Sweet potato badnavirus B). I will continue to use the MiSeq System to perform RNA-Seq and small RNA-Seq.

References

1. Agricultural Research Council, www.arc.agric.za/
2. African Centre for Gene Technologies, www.acgt.co.za/
3. Nhlapo, T. Viral metagenomic analysis of sweet potato using high-throughput deep sequencing. Presented at: 9th Regional Plant Biotechnology Forum; May 22, 2014; Hatfield, Pretoria, South Africa. www.acgt.co.za/wp-content/uploads/2014/06/Thulile-Nhlapo-ACGT-Meeting_22-May-2014.final_.pptx

Learn more about the Illumina system and products mentioned in this article:

- MiSeq System, www.illumina.com/systems/miseq.html
- TruSeq RNA Sample Preparation Kit, www.illumina.com/products/truseq_rna_library_prep_kit_v2.html
- Ribo-Zero rRNA Removal Kit, www.illumina.com/products/ribo-zero-rrna-removal-plant.html

Follow the Illumina Agrigenomics Community

illumına®