

Agriculture Throughout the World is Benefiting from the Latest NGS Technologies

Texas A&M AgriLife uses the NovaSeq™ 6000 and iSeq™ 100 Systems to provide high-quality plant, animal, and insect genome data to its partners worldwide.

Introduction

The human population is increasing rapidly, threatening to outstrip the output of current agriculture practices. The field of genomics is poised to play a pivotal role in enabling farmers to produce enough food to sustain communities worldwide. Next-generation sequencing (NGS) systems and microarrays have been the foundation of agrigenomics research for a decade, providing data to improve crop species and livestock breeds. Charlie Johnson, PhD, the Director of Genomics and Bioinformatics at Texas A&M AgriLife and Executive Director of the Center for Bioinformatics and Genomic Systems Engineering, has years of experience with these technologies. The Texas A&M AgriLife Genomics and Bioinformatics Service lab offers sequencing services to research teams globally to enable studies of a broad range of plant and animal genomes.

The AgriLife lab has been an early adopter of new technologies over the years and serves as a case model for other research groups looking to expand their NGS capabilities. It was the first predominantly agriculture-focused sequencing facility to adopt the NovaSeq 6000 System when it was introduced in 2017. Recently, the AgriLife lab added another NovaSeq 6000 System and an iSeq 100 System to offer additional high-throughput, low cost NGS to its lab.

The AgriLife team is focused on accelerating agrigenomics research that will benefit farmers and support species conservation. iCommunity spoke with Dr. Johnson to learn about his team's work, the benefits his lab has experienced using NGS technology, his recommendations for other groups planning to add NGS systems to their lab, and what the future holds for the field of agrigenomics.

Q: How did you become involved with agrigenomics?

Charles Johnson (CJ): My family has a long history in agriculture starting on a small farm in northern Michigan in the 1800s. In 1894, my great-grandfather, Clarence Beaman Smith, graduated from Michigan State University, which was then the state agricultural college. He went on to work for the US Department of Agriculture (USDA) and was instrumental in the growth of the 4-H association.

I earned a BS degree in soil and crop science at Texas A&M, an MS degree at Clemson University with research in jalapeños, and a PhD degree at Texas A&M University in cotton physiology. I completed a postdoc at the University of Louisville where I worked in computational biology focusing on machine learning and gene

network inference in human systems. My background provided me with great experience and prepared me for my current position where we are sequencing almost every species in the world.

Q: What is the Texas A&M AgriLife Genomics and Bioinformatics Service and what is its mission?

CJ: The AgriLife lab provides NGS and bioinformatics analysis to the research community. At its core, is an outstanding group of scientists with bioinformatics and wet lab experience. Our philosophy is to always work with the latest technologies. The sequencers we have represent more than five different iterations of NGS technology. Our oldest operational sequencer is the HiSeq™ 2500 System, which will be going away soon. We also have HiSeq 4000, MiSeq™, iSeq 100, and two NovaSeq 6000 Systems.

The mission of our lab is to provide high-quality sequencing and bioinformatics for our clients around the globe, and to be the best in the world in that arena. High-quality service is what gets us up in the morning and drives us. We are ever striving to do better and deliver high-quality sequencing results. Working with Illumina sequencing systems makes our mission much easier.

Q: What types of research studies do you undertake?

CJ: We are currently working with research groups in 36 different countries and the range of projects is broad. We perform a wide variety of NGS applications, including gene expression



Charlie Johnson, PhD is the Director of Genomics and Bioinformatics at Texas A&M AgriLife and Executive Director of the Center for Bioinformatics and Genomic Systems Engineering.

studies and *de novo* sequencing, and specialize in using NGS for genotyping. We conduct research studies in most agricultural species, including wheat, cotton, corn, sorghum, rice, and range of vegetables, as well as a host of animal species. Texas A&M has the number one entomology department in the world, so we also do a significant amount of work with insects, including tick, fire ants, kissing bugs, and mosquitos. We work with a broad range of genomic scientists to improve the lives and health of people across the world.

"As we look to sustain a growing human population, we are going to have to use every tool available. NGS is absolutely a part of that."

Q: Why are genomics studies important?

CJ: Agrigenomics studies are important for several reasons. Molecular breeding or marker-assisted selection is a valuable methodology for crop improvement that's better, cheaper, and faster. We are literally translating the instructions of life to understand how to feed the world. As we look to sustain a growing human population, we are going to have to use every tool available. NGS is absolutely a part of that.

Genomics is also critically important for understanding how to conserve wildlife species. Have the populations of certain species already become so small that they're effectively extinct? We can help determine that through genome sequencing.

Q: How did you perform genomic studies before NGS was available?

CJ: I began working in genomics in the late 1990s when microarrays were being developed. That's where I got my start performing gene expression studies and statistical analysis of gene expression data.

The major limitation of microarrays is that they require prior genetic knowledge about a species. You need to know in advance what gene or single nucleotide polymorphisms (SNPs) you are looking for to create a microarray. If you're studying gene expression, you need prior knowledge to select the correct oligos for the array. If the SNPs or oligos aren't on the arrays, they won't be measured.

That's a significant limitation when studying species that have never been sequenced. NGS doesn't require any prior knowledge and that's what makes it such an amazing tool for sequencing plant and animal genomes.

Q: How does NGS enable researchers to decipher the complexities of plant genomes?

CJ: We might never have agriculture reference genomes that match the high quality of the human genome. Most people don't realize that plant genomes are some of the most complex genomes that we study. For example, the wheat genome is five

times larger than the human genome and some pine genomes are twice as large as wheat genomes!

The complexity of plant genomes, such as wheat and corn, made it difficult to design microarrays. It often takes global consortia working together to create the ones that we have. After it is designed, these arrays are inexpensive and easy to analyze and have been tremendously useful in agriculture. However, microarrays aren't feasible for studying many of the species in the world.

As the cost of NGS has decreased over time, it has enabled a level of exploration that was impossible even a few years ago. Now, with NGS, if you want to sequence your favorite species, you can. There is no longer a technical or economic barrier at this point. You can sequence species easily and use the information in your research.

"The major limitation of microarrays is that they require prior genetic knowledge about a species... NGS doesn't require any prior knowledge and that's what makes it such an amazing tool for sequencing plant and animal genomes."

Q: What species are you sequencing with the NovaSeq 6000 System?

CJ: The studies we're performing with the NovaSeq 6000 System are as broad as our client base. We're using it to conduct *de novo* sequencing across a range of species from the smallest bacteria, to insects, animals, and some of the largest plant species. We've done extensive work with agricultural species, as well lions and other large wild cat species, and a wide range of other wildlife and insect species. For example, we just finished projects with five species of tick, ants, kissing bugs, mosquito, deer, quail, feral hogs, prairie chicken, goats, frogs, corn, tomato, wheat, rice, bacteria, cotton, and fungi.

Q: How has the NovaSeq 6000 System transformed your research?

CJ: The NovaSeq 6000 System represents a huge step forward. Our cost of sequencing has dropped significantly. Cost is a significant issue in agriculture and wildlife studies. Before we began working with the NovaSeq 6000 System, it was difficult to conduct large NGS studies due to the higher cost of library prep and sequencing, and the lower quality of the analytics. Today, we can perform much larger studies than ever before, providing more information to breeders who are working with hundreds, if not millions, of different crop lines. With the improved analytics of the NovaSeq 6000 System, we can obtain more information from less data. That enables us to perform low-coverage sequencing, or as we like to call it AgSeq, and sequence hundreds of thousands of samples per year.

Another advantage of the NovaSeq 6000 System is that it's incredibly flexible. Owning a NovaSeq 6000 System is like having eight different sequencers in one system. The S4 flow cell enables us to obtain orders of magnitude more data than with the previous systems. Our last S4 run had 5000 samples on one flow cell. However, we don't always need that much data. We can adjust the output by using the S Prime, S1, or S2 flow cells. We can also gain additional flexibility with the NovaSeq Xp workflow, which enables us to load each flow cell lane individually to separate different projects or methods between lanes.

The NovaSeq S Prime flow cell is a real game changer. For the same cost as a HiSeq 4000 System 150 paired-end run, we can now obtain 250 paired-end reads from a NovaSeq 6000 System run.

The NovaSeq 6000 system is also fast. In the past, a sequencing run would take 4–10 days. It now takes 2.5 days. It's a fantastic machine.

"Today, we can perform much larger studies than ever before, providing more information to breeders who are working with hundreds, if not millions, of different crop lines."

Q: How do you use the iSeq 100 System in your studies?

CJ: Our primary use the iSeq 100 System is for quality control in testing libraries. Quality control is important for high-throughput sequencing runs. Unlike a human genome sequencing lab that might be multiplexing 30–40 samples, we're multiplexing thousands across 10–20 studies. For example, one AgSeq run might have up to 12,000 samples, we have enough barcodes. We use the iSeq 100 and the NovaSeq 6000 Systems in tandem every day. They are a powerful duo.

The iSeq 100 system can also be used to sequence very small genomes, such as bacteria. It's available for those university faculty members who have smaller studies and want a faster turnaround time. We've encouraged them to buy their own iSeq 100 Systems, because it delivers high-quality data, is affordable, and easy to use. You remove it from the box, plug it in, connect it to the Internet, and it will be ready to sequence in less time than it takes to thaw the reagents. With the addition of 250 paired-end reads, it's really going to be a fantastic sequencing tool.

Q: What has been your experience working with Illumina?

CJ: My journey with Illumina began in 2004, when I was part of an FDA microarray quality control project (MAQC), where we evaluated the performance of various microarray technologies. Even then Illumina was known for its quality data and was one of the best platforms. When I started at AgriLife in 2010, we had two Genome Analyzer™_{IIx} Systems. We have upgraded to the latest sequencing technology every 12-18 months. We will be moving

into a multimillion-dollar, state-of-the-art NGS facility in April 2019 that is capable of processing more than 100,000 samples per year. We plan to double this capacity in the future. We'll be sequencing all samples with Illumina NGS systems.

I've always been impressed with the quality of Illumina scientists and the entire operation. In working with Illumina, it feels like we are collaborators more than customers. It's a partnership where both entities are driving toward win-win situations. I truly believe that Illumina has empowered our success over the years.

"Owning a NovaSeq 6000 System is like having eight different sequencers in one system."

Q: What are the emerging applications of NGS?

CJ: As the price comes down and the analytics keep improving, NGS will fuel an explosion of sequencing. We'll be sequencing hundreds of thousands or millions of samples a year. That's per species, so imagine yearly hundred thousand corn samples, hundred thousand cattle samples, or even higher genome sequencing projects. That's the power of this technology. We have met with agribusinesses that are already genotyping millions of samples. I believe that most of them will shift from older technologies to sequencing. If we're identifying a mutation, we ideally need to sequence very large numbers in order to use less data per animal, per plant, etc.

I see NGS as an extremely powerful tool in crop and animal improvement. We're at the beginning of integrating it into the process. It's not like we're just going to sequence each crop species and be done with it. We'll be sequencing every breeder's crop line each year potentially. The idea of being able to sequence every plant/animal species might be a stretch, but there are groups working in that area.

Q: What is automated phenotyping?

CJ: Until recently, phenotyping was performed by going out in the field and measuring plants. Literally, sending out armies of students into fields to walk around with meter sticks and notebooks. The genomics were there, but phenotyping was a road block. The rise of automated technologies, such as unmanned aerial vehicles (UAVs), has changed everything. At Texas A&M, we now regularly use UAVs to fly over fields and take measurements across hundreds of acres in an eight-hour period. That was simply impossible in years past.

Q: What advice would you give a researcher considering making the transition to NGS?

CJ: Just do it. If you're moving into NGS, the first thing to focus on is experimental design. It doesn't matter how much data you have or even the quality of that data. If it's a poorly designed experiment, you're not going to get the data that you need. There's no substitute for quality data and there's no amount of magic or statistics that can make a bad study design good. Next,

find a good service provider, someone you can work with and who you can trust.

"In the future, I can imagine a world where farmers use genomics to identify a fungal pathogen or an insect in the field. The decreasing cost of NGS will enable it."

Q: What would you say to researchers considering using the NovaSeq 6000 System to perform their studies?

CJ: The NovaSeq 6000 System is a technology that will grow with you. You might want to start with the smaller S Prime, S1, or S2 flow cells and branch out to the S4 as necessary. Your lab may not need the S4 now, but you might be interested in a system that will last several years and grow as your needs grow. That's a huge issue, especially in academia where you might only get the funding once.

Q: Where do you see the field of agrigenomics in 5–10 years?

CJ: In 10 years, agrigenomics will be much more common than it is today. In the future, I can imagine a world where farmers use genomics to identify a fungal pathogen or an insect in the field. The decreasing cost of NGS will enable it.

Q: Do you think this is an exciting time to be in genomics?

CJ: Every year I say that this is the most exciting time to be involved in genomics. Recently, I described it as working in a Sci-Fi movie world. The difference is, every year genomics technologies are advancing and studies that were impossible 5–10 years ago, are routine now. For example, the ability to sequence 48 genomes in 48 hours was previously science fiction. Today, we can sequence a sample shipped in a box from the deepest parts of Africa or the Amazon. What is tomorrow going to bring? Who knows, but it's going to be awesome.

See how Texas A&M AgriLife is sequencing complex plant genomes:

Texas A&M AgriLife video, www.illumina.com/company/video-hub/eechnlhAn9w.html

Learn more about the products and systems mentioned in this article:

NovaSeq 6000 System, www.illumina.com/systems/sequencing-platforms/novaseq.html

iSeq 100 System, www.illumina.com/systems/sequencing-platforms/iseq.html

MiSeq System, www.illumina.com/systems/sequencing-platforms/miseq.html

HiSeq 4000 System, www.illumina.com/systems/sequencing-platforms/hiseq-3000-4000.html