



# MARKER-ASSISTED BREEDING

*Accelerating the results of 21st century technology to growers*

Historically, it has taken years or decades to improve crops for such traits as disease tolerance or drought and heat-stress resistance. Today, plant breeders at Texas A&M AgriLife can screen for these complex yet critically important traits in significantly less time, thanks to technology at the Texas A&M AgriLife Genomics and Bioinformatics Service.

The unit, established in 2010, provides next-generation sequencing and bioinformatics services to plant and animal breeders, physiologists and life scientists across The Texas A&M University System. Next-generation sequencing is a technology allowing scientists to decipher the blueprints of all life, which are encoded in DNA. Bioinformatics analysis involves analyzing the enormous amount of data and turning it into usable information.

Since opening, the facility has collaborated with more than 400 scientists and has been involved in more than 190 grant applications, according to Dr. Charles Johnson, director of the service.

Johnson said AgriLife scientists are now able to connect the genotypic, or genetic, information from large numbers of breeding populations to a wealth of phenotypic information, meaning variations in the plant's physical traits and function. This allows them to make the connection between genes and resistance to drought, disease and insects, as well as other high-value traits, which leads to increased yields over a broad range of conditions.

Both phenotypic and genotypic information is needed to understand crop traits.

When working with plant breeders, Johnson said, instead of looking at one or two genetic markers at a time, "we can look at hundreds of thousands of markers at one time. Rather than having to expose a plant to a given stress, we can measure the DNA marker and in a systematic way link that marker to a given trait and use that information to select those plants that carry our traits of interest."

Dr. John Mullet, professor in the Department of Biochemistry and Biophysics at Texas A&M University, said that before advanced technology, acquiring genotypic information was the limiting factor to understanding crop traits. He said getting that information was both very costly and labor intensive.

Mullet said that what one of the first high-throughput sequencing platforms (the Illumina GAI machine) allowed them to do, starting in 2005, was to switch from indirect methods of genotyping to sequence-based genotyping.

"The technology we developed we called digital genotyping because of the AC-GT digital nature of the information," Mullet said.

AC-GT stands for adenine, cytosine, guanine and thymine, key components of DNA.

In addition to genotyping, the center conducts bioinformatics analysis, leading to the selection of molecular markers that become a DNA road map for desirable traits. He compared the bioinformatics analysis to a giant puzzle. A team of bioinformaticians, geneticists, statisticians, mathematicians and computer scientists put the DNA information pieces together to make a complete picture.

Mullet said the bioinformatics analysis is quite intensive. In the past, generating a genetic map consisting of 500 to 1,000 DNA markers took a year. With the first high-throughput sequencing platform, it only took a few months to produce five or six maps, he said. Comparing old maps to new ones showed that accuracy had also increased.

Last year the Genomics and Bioinformatics Service acquired the latest in a series of increasingly fast and powerful DNA sequencing machines: the Illumina HiSeq 2500®.

Johnson said that the top-notch DNA sequencer can complete the equivalent of the human genome project in just 24 hours. The human genome project, which provided the first blueprint for human DNA, took 13 years and cost more than \$2.8 billion by the time it was completed in 2003. Now this same feat can be accomplished for less than \$5,000. ➔





The new sequencer may allow that price to drop to as low as \$1,000, according to Johnson.

“This new technology has been a driving focus in expanding genomic research across Texas A&M,” Johnson said, “and we look forward to expanding the use of this technology across an even wider group of scientists in the future.”

Dr. Bill McCutchen, executive associate director of Texas A&M AgriLife Research, said AgriLife is building an unprecedented understanding of gene content and genome organization. “Quite simply, future discoveries will be all about our ability to measure phenotype in a reproducible and sensitive way and then reduce this information to a description at the molecular level,” he said “Investments in genome science technologies will help move research programs beyond simple information gathering to knowledge generation.”

### Details of the methodology

The researchers’ methodology begins with creating populations that they later segregate according to important traits such as drought tolerance, Mullet said. “Then we score or phenotype those traits under optimum conditions, sometimes in growth chambers and sometimes in the field depending on the trait we’re trying to assay,” Mullet said.

Next, the team identifies the parts of the genome and genes that control the traits. That information enables the team to do marker-assisted breeding and gene discovery, he said, and this methodology streamlines the research process.

“We sort of get right into the nexus there, and we do both gene discovery and marker-assisted breeding,” Mullet said. “What this technology does for our breeding program is it allows us to get rid of things that we know will not work because of past history and lets us focus on things that have the potential to work so we can evaluate more of those. It’s those things then we take to the field and go through our traditional breeding and evaluation.”

### Sorghum sequencing

Dr. Trish Klein, associate professor in the Department of Horticultural Sciences and the Institute for Plant Genomics and Biotechnology at Texas A&M, works with Johnson and with sorghum breeders to develop genetic markers for particular traits, including disease resistance and height, she said.

She agreed that using the gene sequencing technology is tremendously beneficial for the breeding program, saving money and time.

“Using the Illumina HiSeq 2500 platform, we can identify molecular markers linked to traits of interest and then screen sorghum germplasm for these markers,” Klein said. This screening identifies plant lines that contain certain molecular markers and thus are more likely to have a desired trait, for instance resistance to fungal disease.

“This has a tremendous impact on the efficiency of our breeding program,” Klein said.

Sorghum researcher Dr. Bill Rooney, professor in the Department of Soil and Crop Sciences at Texas A&M, said his research group provides the plant tissue. “The Klein and Mullet labs process the plant samples and then the Genomics and Bioinformatics Service runs all the sequencing,” he said. “After Klein’s lab processes the sequence data, she provides information in the form of what sequences are present in which plants.”

Rooney receives a color-coded Excel spreadsheet listing the plant lines with the various markers of interest. Once he receives this information, he selects the best lines to advance in the program based on phenotypic and marker data. Rooney and his group use those genetic associations to shorten the time it takes to do the breeding.

“We can do it faster and eliminate some generations and secondly, maximize the efficiency of our evaluation process,” he said.

For example, Rooney said, if his team has more than 3,000 lines to phenotype, the sequence data allow Rooney and his group to eliminate the genotypes that do not possess the trait instead of having to grow them all in the field.

The Genomics and Bioinformatics Service recently acquired the latest in a series of increasingly fast and powerful DNA sequencing machines: the Illumina HiSeq 2500®. Photo by Danielle Kalisek.

This process reduces the number of plants being tested in the field about ten-fold. Having already selected for genotype, the team can focus on planting sorghum that has the right phenotypic characteristics for the trait of interest.

Starting with the material Rooney or other breeders bring to Klein's group, the group extracts DNA from leaf tissue for each plant sample and prepares it for the HiSeq 2500 machine, Klein said. In the end they may have more than 100 DNA samples in a given tube, each DNA with its own specific barcode. These samples are given to Johnson's group for sequencing. Once the data is returned to Klein, she has a bioinformatics pipeline that can process all the data to identify markers in each of the individual plant DNA samples that were sequenced.

Due to the massive amounts of data, Klein has developed bioinformatic scripts that can find sequence differences among the lines in a largely automated fashion. Her team can correlate those sequence differences to phenotypic differences that have been measured in the field, she said.

"Basically, we use those marker trait associations or genetic mapping studies to identify regions in the sorghum genome that are linked to our traits of interest, and because they are DNA sequence differences, we can turn those into marker assays for rapid screening," Klein said.

The technology Klein's lab uses to make those libraries is called digital genotyping. The particular methodology they use was developed in Mullet's lab.

"Dr. Mullet's group developed the digital genotyping, and my lab developed the pipeline for handling the downstream data," she said.

### Traditional versus sequencing technologies in wheat

While sorghum sequencing is fairly advanced, wheat breeders are just starting to break into the new sequencing technologies.

"I am a very traditional plant breeder, so most of what we do is based on field evaluations," said Dr. Jackie Rudd, AgriLife Research wheat breeder in Amarillo. "Our program develops wheat varieties for producers. Most of what we do is plant yield trials across the state, see which ones come out on top and advance the 'winners.'"

"We're in a natural drought area, and things that survive in multiple field trials here have drought tolerance, no question about it," he said. "Can we improve that? We think so. Can we save time on it and move that drought tolerance a lot quicker than we have in the past? Absolutely, and that's where the

whole genome selection is going to be valuable."

Marker-assisted selection is something Rudd has been doing in his program one gene at a time, he said.

"We developed markers for some wheat rust resistance, green bug resistance and wheat streak mosaic virus resistance; we only have markers for a few specific resistance genes, so our progress is limited," Rudd said. "What really intrigues me is that we can now do whole genome selection. We are just beginning to conceive what the benefit of that is going to be; theoretically it makes a lot of sense and I believe it will change how we do breeding."

Currently genome selection is at the very early stages for the wheat breeder. Particular wheat lines are known to be drought- or disease-resistant or to have bread-baking qualities. Breeders have made multiple crosses with other wheat lines to bring in other useful traits, he said.

"We will use whole genome selection to combine variety A with the best part of variety B," Rudd said. "It is a trial-and-error process through traditional breeding; but this new technology is extremely promising. I'm a little apprehensive since this is our first time to use digital genotyping in wheat, but I'm ready to jump in with both feet."

### Technological advancement

Mullet said he often asks his students, "What technology has been developing the fastest in the last 10 years?" The usual answer is computing technology. But actually, Mullet said, "sequencing technology is improving at a rate faster than computing technology" and is "revolutionizing what we can do."

"This technology is groundbreaking because it's giving biologists a look into DNA in a way that even six years ago was impossible," Johnson said. "It's allowing the biologists and scientists to conduct research that would have been in the realm of science fiction only a few years ago. We are truly part of a new era of life science research."

Rudd is looking forward to the advancements that this new technology could bring to the wheat breeders.

"New technology is talked about all the time," Rudd said. "Every year someone is saying how much better their technique is, but this possibility of whole genome selection really fascinates me. It's based on good science and I'm more excited about it than any other new technology ever in my career, except maybe my first computer."

Read more about the program at [twri.tamu.edu](http://twri.tamu.edu). 