Instrumentation that can sequence the entire genome of an organism is now available on the South Dakota State University campus, thanks to funding from the National Science Foundation, South Dakota Agricultural Experiment Station, SDSU Office of Information Technology and South Dakota’s BioSystems Networks and Translational Research (BioSNTR) Center.

“Having a sequencing center in South Dakota allows researchers to rapidly acquire genomic data, conduct bioinformatics analyses and then refine their experimental designs,” said BioSNTR director Adam Hoppe.

“High-throughput genome sequencing is not an option anymore—this is something researchers have to do to get funded,” said Department of Agronomy, Horticulture and Plant Science associate professor Jose Gonzalez, who led the team that secured funding.

Gonzalez and assistant professor Sunish Sehgal and associate professor Senthil Subramanian as well as biology and microbiology professor Heike Bucking were awarded a three-year, nearly $350,000 NSF equipment grant in fall 2015. The South Dakota Agricultural Experiment Station and the SDSU Office of Information Technology provided an additional 30 percent in matching funds.

Gonzalez oversees the genome sequencing laboratory at the Young Brothers Seed Technology Laboratory, which houses the Illumina NexSeq 500 nucleic acid sequencer along with accompanying instrumentation to prepare samples.

The equipment was selected based on the amount of output that a typical research project needs, he explained. The machine has the capacity to do one or two human genomes in a single run, with an estimated run time of 30 hours.

The lab offers fee-based sequencing services to researchers at universities and biotechnology firms and has a fulltime technician to prepare samples and do the sequencing. Estimates can be requested here.

“We can sequence DNA or RNA,” he explained, noting that there is more demand for the RNA sequencing option. “RNA represents all of the genes that are being expressed at any given point in time and in any given tissue—the working machinery.”

Continued On Page 3...
Determining the order of the DNA nucleotides—adenine, guanine, thymine and cytosine—that make up an organism through next-generation sequencing technologies has become commonplace. However, generating that code of life is only the beginning.

Genome News Network compares a genome to “a book written without capitalization or punctuation, without breaks between words, sentences or paragraphs, and with strings of nonsense letters scattered between and even within sentences.”

To make sense of these big datasets, scientists partner with specialized statisticians who develop algorithms and software that help researchers decipher what’s happening within the code.

That is where the work of assistant professor Qin Ma comes in. He developed software called QUBIC to help analyze RNA gene expression data in 2009 as part of his doctoral work. He completed joint doctorates in operational research from China’s Shandong University School of Mathematics and in computational systems biology from the University of Georgia.

Now he’s updated the software, increasing its efficiency and adding new visualization capabilities. Ma came to South Dakota State University nearly two years ago as a BioSystems Networks and Translational Research (BioSNTR) faculty member in the Department of Agronomy, Horticulture and Plant Science and holds a joint appointment in the mathematics and statistics department.

**Analyzing big datasets**

“He develops tools to help life science people make sense of their data more efficiently,” explained professor Anne Fennell, who has been doing genomic research on cold-hardy grapes for nearly 20 years.

Through a National Science Foundation-funded project, she and Ma will analyze how grafting affects cold-climate grapes. Grapes are commonly grafted, so the root system is genetically different from the top portion that produces the stems, leaves and fruit.

“Plant scientists spend a tremendous amount of time going through these big datasets—anywhere from 2 million to 12 million data points,” Fennell explained. “One of the unique features of QUBIC is that it can effectively analyze these large biological datasets.”

In analyzing RNA from specific segments of the genetic code, scientists seek to determine what the relationships are among the genes: “Some genes are very tightly connected, while others are not,” Fennell pointed out. Traditional clustering looks at the genes under one condition, such as level of expression. However, the relationships are often more complex with patterns changing over time.

Ma’s software uses biclustering to identify whether these genes are co-expressed under varying conditions, such as different temperatures, stresses or daylength.

Continued on Page 4...
2018 Governor’s Giant Vision Competition

The South Dakota Chamber of Commerce and Industry is reminding South Dakota residents who have a new business idea to enter the 2018 Governor’s Giant Vision Business Competition for a chance to win up to $20,000 to launch their idea.

If you’re a South Dakota college/university/technical school student, enter the Giant Vision Student Competition and compete for a top prize of $5,000. High school students are also encouraged to apply.

Complete details are on the website www.southdakotagiantvision.com and entrants will submit their business plan through the website. The competition is open to all South Dakotans with new ideas that they believe would benefit from the review and analysis process of the competition, while vying for funds to assist their start up. The focus of this program is on the business plan or business model and potential for job creation. It is not limited to technology entries.

In addition to cash awards, the competition will bring people with new business ideas together with judges and other invited guests who have business development experience and those who have the capacity to make venture investments.

Application deadlines are in February and the final event competition will be held in conjunction with the annual GOED Conference April 12, 2018 at the Sioux Falls Convention Center.

Sponsors for the Student Competition are Citibank and SD EPSCoR.

DNA, RNA Sequencing

...Continued from page 1

For example, a researcher can grow one plant in normal soil and another in high salinity soil and then look at the differences in gene expression between the two plants. “You might have only a few hundred genes differentially expressed—some are the cause of the difference, while others are the effect of that difference,” he explained. “That gives you a good point from which to start.”

Researchers then combine their findings with other evidence or data to get a better understanding of what’s happening.

- Courtesy of Christi Delfinian
Qubic-R

...Continued from page 2

“Using biclustering, we are getting things that are more similar,” Fennell explained. Ma added, “Identifying biclustering helps elucidate the overall regulatory system and gene expression network.”

Making QUBIC faster

To improve the program’s efficiency, Ma implemented the program using R, a computer language that statisticians and bioinformaticians use. That decreased the average run time by 82 percent compared to the previous version, which was coded in the C language, Ma explained. “Efficiency is very important in the big-data era.”

To test the software, Ma input data containing 900 million elements and processed them in 30 minutes. “This software is very robust—and can be applied to all kinds of data,” added Fennell, including datasets generated by animal scientists, microbiologists and virologists.

QUBIC-R outperformed three popular analysis packages on five RNA datasets—E-coli bacteria, the small flowering Arabidopsis plant, grapes, a human tumor and switchgrass—as detailed in an article the researchers published in the November 2016 issue of Bioinformatics. It is the official journal of the International Society for Computational Biology, the leading professional society for computational biology and bioinformatics.

QUBIC-R is an open access software available free through Bioconductor, one of the nation’s largest bioinformatics companies. “Having it on Bioconductor means that it’s readily available and being used globally,” Fennell said.

Furthermore, Ma pointed out, “You do not have to be bioinformatician to use this program.”

Fennell added, “Even those with not a lot of computational background should be able to easily use the program.”

Visualizing relationships, setting up experiments

QUBIC–R also features enhanced means of visualizing gene expression and looking at specific genes within a network. The software not only creates heat maps which show the level of gene expression based on color intensity in identified clusters, but can also generate coexpression networks that visualize the importance of specific genes within those networks.

Researchers want to know, not only how they are connected, but also how strong those connections are, explained Ma. The software allows the researchers to delete data from specific genes and thereby figure out a gene’s importance in the biological network.

Using co-expression network, he said, “we can see through the size of the circle (that represents a specific gene) when more interactions are taking place. The thickness of the lines (or connections) also gives you more of the dynamics.”

Experiments are designed to test a hypothesis. Therefore, researchers tend to focus on parameters related to their hypotheses. By using QUBIC-R as a computational tool, researchers can do “hypothesis-free” data exploration.

“This type of computational analysis can show you what you might be missing because of that hypothesis-driven focus,” Fennell said.

Through his role in BioSNTR, Ma helps researchers design their experiments, determining how many samples to prepare and what kind of sequencing is suitable to test the hypothesis. “It’s very important to think computationally when you are developing your experiments,” Ma said.

Fennell added, “Qin can suggest tweaks in experimental design that will give you better analysis in the end. BioSNTR is about teams coming together to solve a problem by identifying things we weren’t able to do before.” Bioinformatics and the analytical tools that Ma develops help make this possible.

- Courtesy of SDSU Marketing & Communications
Science Discovery Day

Since spring of 2011 Sanford Research has hosted over 2700 high school student at the semi-annual *Science Discovery Day* event. Each fall and spring they invite over 250 students to learn about the passion, experience, and current projects in research and biomedical science in the region. The next Science Discovery Day will be held **March 28**.

Sanford invites high school sophomores and juniors with a genuine interest in the sciences to join us and get exposure to hands-on techniques and skills utilized by regional science professions. The program includes career presentations, hands-on activities, and exhibits from area industries and universities. Attendees will receive lunch and a Sanford PROMISE T-shirt. Presenters include professionals from Sanford Research and other area biomedical industry and academic institutions. Click [here](#) for more information on attending the next Science Discovery Day.

Sanford Research also invites research professional to participate by hosting an interactive workshop session or exhibitor booth that highlights the science and research happening at your school or organization.

Links of Interest:

- [Scientists Discover How to Talk to Kids](#)
- [South Dakota Wins Competitive Grant to Support Bioscience Growth](#)
- [New SD Mines President Shares Goals](#)
- [Suspectsyn Named SD State's Director of Institutional Research](#)
- [SD Mines Adds Ph.D. Specialization in Mining Engineering](#)
- [Summer Program for Undergraduate Research in Addiction](#)
- [Search for NSF Research Experience for Undergraduate's (REU's)](#)
- [Breakthrough Biotech Video Scholarship Competition](#)

Important Dates:

- **Deadline to Submit Biotech Video Scholarship Entries** - Jan. 31, 2018
- **26th Annual SDCTM & SDSTA Joint Conference** - Feb. 8-10, 2018
- **Sanford Promise Community Lecture Series: Rare Disease Day** - Feb. 22, 2018
- **Governor's Giant Vision Student Competition** - Application Deadline: Feb. 23, 2018
- **Student Poster Session at the Capitol** - Feb. 26, 2018
- **Coalition of EPSCoR/IDEA States 2018 Annual Conference** - Feb. 28 - March 1, 2018
- **RII Track 4 Proposals Due** - March 13, 2018

This material is based upon work supported by the National Science Foundation/EPSCoR Grant Number IIA-1355423 and by the State of South Dakota. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.