

INNOVATION CENTRE DELIVERS AFFORDABLE, COMPLETE MICROBIAL GENOMES WITH SMRT® SEQUENCING

Scientists at the sequencing core facility in the McGill University and Génome Québec Innovation Centre have raised the bar on assembly performance and read length at an affordable price using their PacBio® RS II Sequencing System.

At the McGill University and Génome Québec Innovation Centre, many projects conducted in the sequencing core facility fall under the umbrella of life sciences rather than biomedical studies. To the scientists responsible for making the core facility operate as smoothly as possible, that makes a world of difference.

“When you’re in the life sciences in addition to human biomedical, you’re out there in the world of things that haven’t been sequenced before, or haven’t been sequenced particularly well,” says Ken Dewar, a principal investigator at the Innovation Centre. “It doesn’t take long before you scratch the surface and realize you’re seeing things that have never been seen in science before. There’s nothing in the databases as references. You’re completely on the dark side of the moon.”

To navigate this type of uncharted territory, scientists at the center rely on long-read sequencing from their PacBio platform to cost-effectively close microbial genomes, traverse repeat-heavy genomic regions, and perform full-length transcript sequencing.

The Innovation Centre has an unusual advantage among core facilities: its size. “Instead of having small cores at every university, there was more of an emphasis to build a larger genomics core for all of eastern Canada,” Dewar says. “It’s a partnership between the university and the



province of Québec.” This gives the scientific team at the center exposure to a broad range of projects and the opportunity to develop expertise across many types of experiments rather than specializing in a narrower scope.

Using that expertise, the center’s scientists have made significant strides with Single Molecule, Real-Time (SMRT) Sequencing. By leveraging the dramatically increased read lengths the PacBio RS II provides, they have driven down costs and improved completeness of their assemblies.

The Innovative Core

Scientists take the “innovation” part of their facility’s name seriously. At the center, which originally began services at a local hospital in 1996 but moved to its dedicated building in 2003, there is a large team devoted to the technology service platform. Additionally, a number of McGill-affiliated faculty members and their labs are located at the center and are responsible for contributing to the technology operations. Dewar, for example, is an Associate Professor in the Department of Human Genetics at McGill, but his lab is based at the Innovation Centre because one of his goals is developing technology protocols. “What we do involves earlier access to technologies, so when we achieve something useful, that gets transferred back to the service platform,” he says.

While Dewar keeps a close eye on the core technologies, it is Alexandre Montpetit who is dedicated to running the next-generation sequencing platforms at the core facility. His primary affiliation is with Génome Québec, and he has an adjunct appointment at McGill. Montpetit oversees a sizable inventory of next-generation sequencing platforms, including Illumina® HiSeq® and MiSeq®, Ion Torrent® PGM, Roche® GS FLX, and PacBio® RS II.

The center’s team of over 40 managers, technicians, and bioinformaticians tackle a wide range of scientific projects.



Facility name:	McGill University and Génome Québec Innovation Centre
Staff:	100-plus at the Centre, with more than 40 managers, technicians, and bioinformaticians dedicated to the core facility and another 60 to 70 McGill scientists who routinely contribute to the technology platform
Year founded:	1996, moved to its dedicated building in 2003
Serves:	900 investigators per year
PacBio System installed:	December 2011, upgraded in August 2013 to PacBio RS II
Website:	http://gqinnovationcenter.com

The most common categories include human genetics, agriculture and forestry, and microbial studies, Montpetit says. The core's affiliation with Genome Canada means that it performs a lot of work for investigators funded by the government agency, and those projects often include sequencing or genotyping trees, fungi, yeast, cattle, pigs, and more.

With about a dozen bioinformaticians on staff, the core facility differentiates itself with the breadth of analysis options it is able to offer clients. "When we started sequencing microbial genomes with 454, everybody wanted the reads," Dewar says. "Then we started offering the assemblies and nobody wanted the reads anymore — they go straight to the assemblies. We see it happening even more with PacBio. We'd rather provide finished genomes than read sets." Now, the Innovation Centre team has raised the bar again: for some projects they produce a custom browser, preloaded with the client's genome assembly, already annotated for various sequence characteristics, genes, and more, Dewar notes.

The SMRT Approach

Montpetit and his colleagues have been champions of long-read sequencing for years, so when PacBio unveiled its platform with industry-leading read length, it was an obvious choice for the center to adopt the technology.

"We've always had a focus on sequencing things for the first time or assembling genomes for the first time, not for the thousand-and-first time," Dewar says. "PacBio was a natural fit." Acquired with a generous donation from the Bachynski Family Foundation, the sequencer was delivered in December of 2011, and demand has risen steadily for it ever since. "Having support from philanthropic organizations really allowed us to get much earlier access to the technology," Dewar says.

At the center, SMRT Sequencing is most commonly used for high-quality assemblies in microbial sequencing, and is also deployed to cover long repetitive genomic regions or to sequence full-length human gene isoforms. Microbial sequencing encompasses a number of applications, including biotech industry efforts to improve microbial biofermentation and microbiome studies, from environmental remediation projects on Alberta tar sands to veterinary health studies of cattle rumen.

In the two years they've been running the SMRT Sequencing platform, the Innovation Centre scientists have seen remarkable progress in what they have been able to achieve. Continued improvements in read lengths — partly due to new reagent kits from PacBio and partly due to more streamlined sample prep protocols developed at the center — have already made a major difference. For instance, at the end of 2013, scientists were getting more reads in the ≥ 11 Kb size range than the reads ≥ 3 Kb that they were getting just six months earlier. "We spend a lot of time emphasizing the importance of really good, high molecular weight preps," Dewar says of the sample preparation improvements.

One major step was achieving complete bacterial sequencing and assembly in less than a day, a feat that may enable the core facility to serve as a rapid response center for organizations that need to tackle pathogen outbreaks and other urgent problems. In 2013, tests conducted with researchers at the Canadian Food Inspection Agency and other government agencies demonstrated that the Innovation Centre scientists could sequence a sample and fully assemble the genome and plasmid elements — all in 20 hours or less.

As the scientific team gained expertise with SMRT Sequencing, they shifted from relying on hybrid genome assemblies to a PacBio-only strategy.

"Six months ago, our recommendation to customers was 100x of MiSeq data with 30x of PacBio data," Montpetit says. "Now we tell them we can do a better assembly with PacBio-only data."

In fact, the success of PacBio-only assemblies using HGAP and Quiver has made scientists at the Innovation Centre reluctant to perform hybrid assemblies. Dewar says the team will only create a hybrid assembly now if the client already has a short-read data set from the same organism.

"The number of corrections we're doing on the PacBio assembly is less than one per megabase, so in most cases it's not worth paying the money to do the Illumina run," he adds.

With these assemblies, the Innovation Centre team is routinely able to deliver affordable, high-quality, finished genomes. "A single bacterial genome, a library prep, and two SMRT Cells of sequencing — which is generally a little bit overkill — is less than \$1,000," Dewar says. "More and more often, we are getting a completely closed, finished-quality genome for that."

High-quality assemblies aren't just for bacteria. "We've shown recently that we can assemble a fungal genome of 20 or 30 megabases with four or eight SMRT Cells and get only 10 or 20 contigs — which often represents the number of chromosomes in the genome," Montpetit says. Projects like these are the reason his PacBio platform has been busier than ever lately: between positive feedback and word of mouth, more and more scientists are coming to the center to use the long-read technology. "A lot of them want to assemble a bacterial genome, and now PacBio is the first solution we offer," he adds. "We're confident because of the successes we've had. The next wave, which we are beginning to work on, is of microbiomes and other complex mixtures."