Taking Genomics to the Next Level

At SciLifeLab, National Bioinformatics Infrastructure Sweden (NBIS) at Uppsala University, Björn Nystedt and his colleagues utilize the most advanced sequencing technology to the fullest in advanced genomic research and diagnostics. For this, PacBio's long-read sequencing instruments play a crucial role, as long-read sequencing gives answers to questions that short reads simply cannot.

By Rikke Rosenlund

A house of technology

Björn Nystedt, Head of Bioinformatics Long-term Support at SciLifeLab, is a bioinformatician who has spent his career working on genome assemblies in order to understand evolution and biology. Nystedt has worked with several different sequencing platforms and technologies during his time with SciLifeLab, a core facility serving all of Sweden with bioinformatics. Today, at the outset of 2020, SciLifeLab continues to house the most innovative technology and holds two PacBio Sequel I instruments and now also a Sequel II.

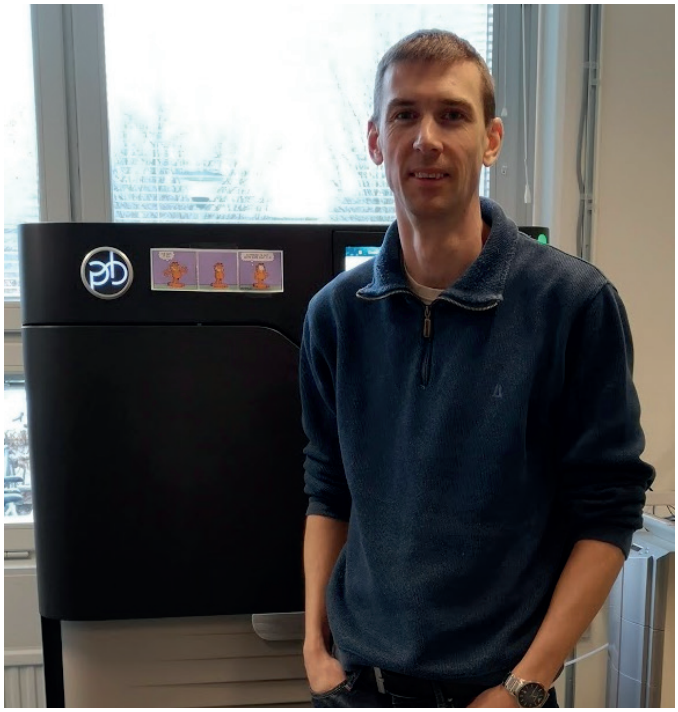
A big leap forward in conifer forest genomics

Norway spruce is a good example of a very large genome that needed a better reference genome. That was what Nystedt and his colleagues at SciLifeLab set out to tackle in 2013, when they sequenced the 20-gigabase genome of Norway spruce. At the time, it was the largest genome ever sequenced. But despite using the best technology available in 2013, "the project had a lot of limitations because the genome assembly was very fragmented," tells Nystedt. What is especially interesting about the Norway spruce is its genome size combined with an inadequate reference genome and a massive amount of retrotransposable elements, in terms making it a difficult task to de novo assemble the genome.

> **"The benefit of longer reads is obvious in so many settings"**

Currently, Nystedt and his team are, among other projects, working on the second round of the spruce and pine project, using PacBio's Sequel I and II for data generation: "We are trying to take conifer genomics to the next level. We took a huge leap forward in 2013, and this will be a big leap forward again. This project will have large implications," says Nystedt.

But why is the spruce genome so interesting? Nystedt explains: "There is almost no completely natural forest in Sweden, as almost all Swedish forestland is managed and utilized for commercial forestry. One aim of the project is to approach breeding in a more scientific and sustainable way." A good breeding program is important as conifer makes up more than 80 % of Sweden's forests. PacBio's Sequel instruments and the long-read sequencing technology have turned out to completely remodel the work at the platform and the instruments are therefore crucial in the work to answer complex genomic questions. In addition to conifer genomics, the team of researchers have also done some interesting development on clinical applications on PacBio, Nystedt adds.



A more comprehensive view of genomes

Over the past decade, we have seen great advances in next-generation sequencing technology, and the long-read sequencing technology has changed the way the team works on the spruce project. The spruce project from 2013 was based only on short-reads, whereas the team currently makes use of long-reads. "The pieces of DNA that we had at the end of the last project, after all the assembly and work we did, they were still shorter than the reads we start with in this project, before we do any assembly," says Nystedt, and continues: "We are aiming for pieces that are 3-4 magnitudes longer than in the last project. It makes a huge difference when understanding the genetic context."

> **"The Sequel has been a game changer for us. A bacterial genome literally comes out perfect from the machine"**

The benefits of long-read technology are hard to miss in many applications, according to Nystedt. He highlights that the time it takes to run an assembly goes down, but more importantly, that there is much less risk of errors when doing the assembly – "It is about the quality of your work," he says. In addition, when computing the long-read sequences, it is mathematically a simpler problem because you do not have to account for all these errors.

The ability to generate highly accurate long reads (HiFi reads) is a paradigm shift: "I am probably one of the few bioinformaticians who have run a Sanger sequencing machine myself, and to now have high-quality reads of 20 kbp is extremely impressive! In a lot of ways, it is a dream come true because with previous long-read technology, high error rates were the price you had to pay. Now, that is shifting because of the ability to generate that type of high-quality data," says Nystedt.

New possibilities arising from fast-paced technological progress

SciLifeLab has been at the forefront in taking on the evolving technology. Nystedt therefore has firsthand experience with sequencing data from short-read platforms, the PacBio RS ii system as well as the PacBio Sequel I and Sequel II systems. The progress from PacBio has indeed been tangible, according to Nystedt: "We have seen major improvements in every upgrade so far. PacBio has for us been the leading long-read technology – it has been the major workhorse for us when it comes to long reads."

Most whole genome sequencing projects that have been run at SciLifeLab in the past 2-3 years have been run at the Sequel I instrument. "The Sequel has been a game changer for us," says Nystedt, and explains that it has transformed the way they work because the machine can sequence genomes perfectly: "A bacterial genome literally comes out perfect from the machine." Previously, they would spend massive amounts of time to close these genomes. The system also opens for new research possibilities, he adds: "We will now be able to explore much more complex events than what we used to."

> **"We have seen major improvements in every upgrade so far. PacBio has for us been the leading long-read technology - it has been the major workhorse for us when it comes to long reads"**

SciLifeLab had no hesitations in upgrading to the most recent technology by adding a Sequel II system to their laboratory. Nystedt explains: "It seems that technology is continuing to develop fast. We have just had our Sequel II installed and from what we have seen in our first runs, it is again a major upgrade. It is clear!"

Combining technologies to get the full picture

The long-read sequencing technology has proven to be extremely useful to the ongoing studies of conifer, but Nystedt does not shy away from using short-reads as well: "We are using a lot of different technologies in our work," he says. He emphasizes how the long reads are essential to address many research questions and by using complementing technologies, he can get a fuller picture.

The vast potential of long-read technology

The vast potential of longer reads in sequencing is crystal clear to Nystedt: "The benefit of longer reads is obvious in so many settings," he says. He points to some major areas that may change in the future: "We begin to see these global collaborations where you are starting to sequence basically all life on earth, which lets us finally know more – there are super fundamental questions we could get answers to. As well as with virus outbreaks and pests, and then there will be many clinical applications where you just need to sequence longer stretches to get the full picture. These technologies will enable that."

According to Nystedt, one of the challenges as technology progresses is that we generate much data that goes unused. "To take advantage of all the possibilities we have now, the community needs to invest even more not only to generate good data but to also have the capacity and the competence to analyze and understand all the different aspects of a project, because there is fantastic technology to build on," Nystedt says.

Sequence with Confidence

PACBIO®

PacBio Sequel II

The Sequel II System has been recognized for its ability to generate longer reads with greater accuracy and throughput, at a significantly lower cost than other systems.

The system gives all scientists access to high-throughput, cost effective, highly accurate long read sequencing that can solve the unsolvable by short read sequencing:

- Whole-genome sequencing with allele-specific haplotypes
- RNA sequencing with isoform-level resolution
- Targeted sequencing through low-complexity regions
- Complex populations, identify closely-related individuals within a heterogeneous mixture
- Uncover the role of epigenetics in the inheritance of traits from one generation to the next

2019 TheScientist TOP 10 INNOVATIONS

The Life Science LSIA 2019 Industry Awards

AH diagnostics

AH diagnostics A/S • Gammel Viborgvej 11A • DK-8381 Tilst • Tel.: +45 8745 9010 • ahdiagnostics@ahdiagnostics.dk • www.ahdiagnostics.dk
AH diagnostics AB • Vallgatan 9 • SE-170 67 Solna • Tel.: +46 (0)8 680 0845 • ahdiagnostics@ahdiagnostics.se • www.ahdiagnostics.se
AH diagnostics AS • Postboks 4499 Nydalen • NO-0403 Oslo • Tel.: +47 2323 3260 • ahdiagnostics@ahdiagnostics.no • www.ahdiagnostics.no
AH diagnostics Oy • Viikinkaari 4 • FI-00790 Helsinki • Tel.: +358 (0)10 325 3000 • ahdiagnostics@ahdiagnostics.fi • www.ahdiagnostics.fi