

IN THE NETHERLANDS, SCIENTISTS CRACK TOUGH PLANT GENOMES WITH PACBIO® SEQUENCER

At KeyGene, agricultural specialists strive to improve plant varieties through molecular breeding and other advances. They rely on SMRT® Sequencing to produce top-notch assemblies from even the most difficult genomes.

When it comes to battling genome complexity, the scientists at KeyGene are veterans. The Dutch company focuses on innovations in plant breeding, so its researchers have a mandate to decipher the genomes of some of the most complex organisms. Plants contain the difficult elements that plague all genomic scientists — such as highly repetitive sequences, GC-rich content, and structural variants — with the added challenge of polyploidy thrown into the mix.

In an industry where many companies survive just a few years, KeyGene is celebrating its 25th anniversary and has more than 140 employees. That longevity is based on a successful track record of using advanced breeding technologies and trait improvement approaches to assist with crop development. “Our mission is to develop innovative technologies to accelerate plant improvement,” says Michiel van Eijk, the company’s chief scientific officer. “The challenge is to develop technologies which are broadly applicable in as many species as possible.”

To meet that challenge, KeyGene scientists closely monitor DNA analysis technologies and are often early adopters of new tools they believe will lead to better plant characterization. They brought next-generation sequencing technology in-house as soon as it was commercially available. When Pacific Biosciences introduced its Single Molecule, Real-Time (SMRT®) Sequencing, KeyGene researchers were eager to try out the world’s first long-read technology. Today, the PacBio® sequencer has become an integral part of the KeyGene pipeline, delivering average read lengths of 10,000 – 15,000 base pairs. “When it comes to genome assembly, PacBio is our first choice,” van Eijk says.

The KeyGene Vision

Formed in 1989 by several Dutch seed companies, KeyGene has established its reputation as a leader in the science of understanding plants. “We develop and apply technologies to help plant breeders improve their varieties,” van Eijk says. “That was the vision with which KeyGene was founded.”

KeyGene works with its founding companies on vegetable species, and with customers around the world for field crops, commodity crops, and other plant species. Many of



Scientists at KeyGene depend on superior genome assemblies to explore the genetic basis for important crop traits, working with breeding companies to develop new varieties that are more productive, resistant to disease, and better adapted to the changing environment.

these organisms have genomes that are quite large and feature marked complexity. Wheat, for instance, has an estimated genome size of 16 Gb.

In the early '90s the team introduced its AFLP® technology, which was quickly embraced by the community for detecting genetic polymorphisms. Originally performed with PCR and gels, the AFLP technique’s utility across all plant species provided the impetus to launch a service business, which continues at KeyGene today. Customers send in samples “for standardized analyses like genotyping, mutation screening, or sequencing,” van Eijk says.

To unravel difficult genomes, van Eijk and his team rely on cutting-edge tools. “We are continuously on the lookout for technologies that can help us achieve innovations in molecular plant breeding,” he says. Next-generation sequencing, which van Eijk first encountered in 2004, is one example. “The first time I realized there was technology capable of producing hundreds of thousands of sequences, the connection with our own AFLP technology became clear immediately,” he adds. At the time, AFLP was limited by the need to resolve fragments with gel electrophoresis. KeyGene was an early adopter of the first next-gen sequencing technology, the 454 platform, and

has been a champion of cutting-edge sequencers ever since.

“That jump-started the thinking that we were entering an era where large-scale sequencing was becoming a reality,” van Eijk remembers. Ever since, his team has been looking for new approaches to DNA sequencing that address issues left unresolved by earlier generations of sequencers, such as systemic error modes and short reads that can’t span complex regions.

That quest led them to SMRT Sequencing and its uniquely long reads. KeyGene acquired the PacBio sequencer in 2011, the first year the system hit the market. With its high accuracy and industry-leading read lengths, the platform has replaced the 454 as KeyGene’s workhorse for generating high-quality genome assemblies, both for collaborative projects and for its service business.

Better Genome Assemblies

With long sequence reads, the KeyGene team is able to dramatically improve plant genome sequences that have been generated with other technologies — in one case by more than a 40-fold boost in assembly quality. “For most important crops, a draft sequence is available, but the quality is very much variable,” van Eijk says. “A draft genome still means there are many gaps and often that the contigs are very small.”

The deep dive KeyGene scientists take into plant biology means that they need more complete, polished assemblies. “The value of a long-read platform is that much better assembly metrics can be reached,” van Eijk adds. “This translates into significantly faster and more accurate gene identification and accelerates trait improvement for breeding.”

The team began evaluating SMRT Sequencing by analyzing the melon genome, which was the first plant they sequenced with their 454 platform. Their existing genome sequence had a contig N50 of about 10 kb. “That’s useful to some extent, but if you want to identify genes based on markers, then contigs of that size are too small to efficiently isolate a gene,” van Eijk says.

“It’s fairly well known
that for genome
assembly, the PacBio
platform is superior”

He and his colleagues sequenced the same melon variety with the PacBio RS II. At first, they planned to use the PacBio data to patch the 454-generated genome. But after the first pass with SMRT Sequencing — generating 20x coverage of the melon genome — the scientists quickly realized that the PacBio-only assembly was already better than the 25x draft genome they were trying to upgrade. “It started out as a gap-closing activity,” van Eijk says. “But it became clear that for *de novo* genome sequencing, it’s much easier and actually yields better specs when you do the assembly based on the PacBio data alone.” The scientists did some more sequencing, upping the PacBio coverage to 44x. They wound up with a *de novo* assembly that was far more complete and had a contig N50 of nearly 400 Kb — a 40-fold improvement over the initial draft assembly. “That’s a much more valuable tool for doing gene identification,” van Eijk says.

Since that initial effort, KeyGene has ramped up its use of the PacBio platform, increasing the size and

complexity of genomes sequenced with it. The team’s latest project represents some of the toughest characteristics: the tetraploid cotton genome, with a genome size of 2.5 Gb.

“It’s fairly well known that for genome assembly, the PacBio platform is superior,” van Eijk says.

The KeyGene R&D team isn’t limiting its sequencing efforts to plants; the PacBio platform is now widely incorporated into sequencing projects for plant pathogens, as well. “We have seen examples where we’ve sequenced these pathogens to closure, into a single contig,” van Eijk says. The researchers also sequence BAC clones and have recently demonstrated the ability to pool 96 clones and sequence them together.

They are continuing to explore new applications of the platform. KeyGene scientists are using the PacBio RS II for transcriptome sequencing with the Iso-Seq method, and are working on reduced-complexity protocols for base modification detection in a subset of the genome. Clearly, the team is making the most of its long-read sequencing system. “The volume of projects keeps increasing,” says van Eijk, noting that the platform is popular for internal projects as well as among customers in agricultural research from all over the world.

www.pacb.com/pag

The AFLP® technology is protected by US patent 6,045,994 owned by Keygene N.V. AFLP is a registered trademark of Keygene N.V. Other (registered) trademarks are the property of their respective owners.

For Research Use Only. Not for use in diagnostic procedures. © Copyright 2015, Pacific Biosciences of California, Inc. All rights reserved. Information in this document is subject to change without notice. Pacific Biosciences assumes no responsibility for any errors or omissions in this document. Certain notices, terms, conditions and/or use restrictions may pertain to your use of Pacific Biosciences products and/or third party products. Please refer to the applicable Pacific Biosciences Terms and Conditions of Sale and to the applicable license terms at <http://www.pacificbiosciences.com/licenses.html>.

Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell and Iso-Seq are trademarks of Pacific Biosciences in the United States and/or certain other countries. All other trademarks are the sole property of their respective owners.