



Keygene Offers Sequence-Based Physical Mapping Service to Improve Plant Genome Assemblies

February 01, 2011

Keygene Offers Sequence-Based Physical Mapping Service to Improve Plant Genome Assemblies

By [Julia Karow](#)

Agricultural genomics company Keygene has developed a method to generate sequence-based physical maps of genomes, which it offers as a service to help assemble a variety of plant genomes.

The company has used the proprietary method, called Whole Genome Profiling, to complete more than a dozen plant genomes and is working on additional genomes for academic and industrial customers. The company is offering it as a service with its co-marketing partner, Amplicon Express. A description of the method was recently accepted for publication in a peer-reviewed journal.

Whole Genome Profiling generates a profile of short sequences tags for each BAC in a BAC library. Keygene researchers pool the BACs in a multi-dimensional format, cut the DNA with restriction enzymes, and sequence the ends of the fragments using the Illumina Genome Analyzer. They then assign the sequence reads to individual BACs wherever possible and assemble them into contigs, and a physical map, based on sequence information from overlapping BACs.

According to Michiel van Eijk, vice president for upstream research at Keygene, the approach is similar in principle to the SNaPshot method for generating physical maps from BAC libraries, which he said has been the "gold standard" but provides no sequence information.

Whole Genome Profiling, he said, "brings the same concept to higher resolution," and the sequence tags can serve as anchor points for *de novo* genome assemblies.

With a typical size of 125 kilobases, each BAC clone is more than five times larger than the distance spanned by the largest mate pairs for next-gen sequencing — about 10 to 20 kilobases — he said. "So the ability of the sequence-based physical map to obtain an assembly of large scaffolds is much higher" than with mate-pair and paired-end reads alone.

On a per-clone base, he said, the cost of a physical map generated by Whole Genome Profiling is comparable to that of a SNaPshot-based map. "But the added value is that you have the sequence-based information," which for a SNaPshot map would require additional BAC end-sequencing.

He declined to provide pricing information for the service, which he said depends on the number of clones to be analyzed, which in turn is a function of the depth of the BAC library and the size of the genome.

Because of the high cost of clone-based libraries, researchers have been trying to move away from them for sequencing plant and animal genomes *de novo*, but according to van Eijk, "at this point in time, it's still a necessity because alternative technologies do not provide us with the possibility to do mate-pair sequencing over distances of over 100 kilobases. If indeed the technology would be there that enables this, then the clones would probably no longer be necessary."

He said that although some genomes have been sequenced solely by 454 or Illumina sequencing at very high coverage, the metrics of those assemblies are worse than if a physical map is added to the mix. "Even if you sequence 100x deep, you still have many more contigs and scaffolds than when you integrate that with Whole Genome Profiling," he said.

Also, with a clone library, researchers have access to BAC clones for any region of the genome of interest, for example in order to clone specific genes.

"While the price of *de novo* sequencing is going down, at the end of the day, you also have to look at what the metrics of the assembly are that you then get, and how well you can utilize that for the purpose that you generated the sequence in the first place," van Eijk said. "And I think now a number of people realize that producing the sequences is relatively cost-effective, but to make a high-quality reliable assembly out of that is something else."

[pagebreak]

Whole Genome Profiling can be applied to any genome for which BAC libraries can be generated, he said, and Keygene has used it to assemble genomes of different sizes, including the 4.5-gigabase tetraploid tobacco genome.

Rod Wing, a professor of plant sciences at the Arizona Genomics Institute of the University of Arizona, said that data he has seen for the Keygene method "looks very good," although he has not used the method yet.

"I am still a strong proponent of using physical maps as backbones for the generation of new reference genomes," he said, noting that several genomes published without the use of physical maps have more than half the genome missing. "I feel it is extremely important that the first

genome to be generated from most organisms should be as high quality as possible. Once this is achieved, then one can resequence till the cows come home." Optical maps, he added, are also a good tool to check the fidelity of a genome assembly.

According to Greg May, president of the National Center for Genome Resources, Keygene's method offers a physical map in a shorter amount of time and at a lower cost compared to other methods. He also thinks that physical maps will have some staying power. "It will be a very long time before next-gen reads only are sufficient to generate a finished genome," he said, especially in polyploidy plants. "Physical maps go a long way in building bigger scaffolds."

And even researchers that try to avoid them agree that they are still useful. Kevin Folta, a professor of horticultural sciences at the University of Florida in Gainesville, recently published the strawberry genome using only next-gen sequencing and a "very saturated" linkage map. This and the fact that the genome is small and only has a modest amount of repeats meant that he and his colleagues could do without a physical map. But "for larger genomes, I could imagine that some level of BAC support or optical maps would be useful," he said.

"There is still a place in the business for it for sure," he said, although he admitted that he has avoided techniques like BAC-based physical maps "like the plague."

Have topics you'd like to see covered in *In Sequence*? E-mail the editor at [jkarow \[at\] genomeweb \[.\] com](mailto:jkarow@genomeweb.com).

Related Stories

- [Illumina Sequencing Revenue Grows 70 Percent in Q4, Driven by HiSeq](#) 🗝
February 10, 2011 / [In Sequence](#)
- [Illumina and GenoLogics Will Co-Market Next-Gen Sequencing Informatics Software](#) 🗝
February 8, 2011 / [In Sequence](#)
- [Wash U Researchers Embark on Targeted Sequencing Project of 450 Childhood Leukemia Samples](#) 🗝
February 1, 2011 / [In Sequence](#)
- [Illumina Opens Office in Brazil to Target 'Increasing' Latin American Market](#) 🗝
January 25, 2011 / [BioArray News](#)
- [In Sequence Annual Survey: Illumina Leads Market but Most Users Believe PacBio Will Provide Next Big Leap](#) 🗝
January 25, 2011 / [In Sequence](#)

