



Longas Technologies Pairs Chemistry, Bioinformatics to Combine Short Reads into Longer Ones

May 22, 2019 | [Andrew P. Han](#)

NEW YORK (GenomeWeb) – Last week, Australian startup Longas Technologies emerged from years in stealth mode, headed by a familiar face: former Solexa executive Nick McCooke.

Now, the firm has revealed details about its proposed product, Morphoseq, a library preparation kit that, when paired with the firm's proprietary bioinformatics package, creates "virtual long reads" from short-read sequencing data.

On Tuesday, Longas unveiled Morphoseq at the Sequencing, Finishing, and Analysis in the Future meeting, held in Santa Fe, New Mexico. Prior to the presentation, Longas shared details about the product and its plans to bring it to market.

Morphoseq is designed to run with any short-read sequencing platform and feed into any bioinformatics pipeline, said Aaron Darling, Longas cofounder and chief scientific officer. Darling is also a professor at University of Technology Sydney (UTS.) Prior to fragmentation, the Morphoseq chemistry introduces uniformly random mutations into longer DNA templates. Once sequenced, the reads can be mapped back to those templates, creating the so-called "virtual long read," which can be up to 10 kb.

The technology behind Longas' product is likely less recognizable than the company's leadership. Sequencing analysis by mutagenesis (SAM) has been around since 2004 and was proposed as a library preparation method for next-generation sequencing in 2012. But unlike Solexa's sequencing-by-synthesis, SAM never really got traction. Longas' bet is that years later, it will.

"The overall idea of SAM is to eliminate problematic sub-sequences via mutation, sequence a number of such mutants, then infer what the original sequence must have been," said Jonathan Keith, a professor at Australia's Monash University. In 2004, Keith was part of a team that [published SAM](#) as a method for eliminating problematic sequences using Sanger sequencing.

"It's cool someone's trying it," said Nick Goldman, joint head of research at the European Molecular Biology Laboratory-European Bioinformatics Institute. In 2012, Goldman led a team of researchers that included current Oxford Nanopore Bioinformatician Botond Sipos who [described a theoretical protocol](#) for

getting SAM to work with NGS. However, their method — which at a high level resembles Morphoseq — was only tested *in silico*, and not in a wet lab, Goldman said.

"Even if all they did was optimize [SAM] to something that works reliably in the lab, well done to them. That would not be trivial to do," Goldman said.

The company joins 10x Genomics and BGI in the market for technologies that approximate the read lengths of single-molecule sequencers using short-read platforms. Unlike 10x's linked read sequencing, it does not require a separate instrument. And unlike BGI's long read fragment technology, Morphoseq is compatible with a number of sequencing technologies.

"It basically overcomes the Achilles' heel of short-read sequencing, which is the [in]ability to get read lengths necessary for genetic challenges," McCooke said. Joining the firm as CEO "was a nice way to come full circle and help to remedy one of the things we didn't get right [at Solexa.]"

McCooke was conservative with his predictions for the value Longas could bring to the industry. "We're not making an instrument. It would be a little silly to say it can be [the kind of opportunity Solexa was,]" he said. "But we want to have impact. We're highly motivated."

Longas, [a spin out from the itthree Institute at UTS](#), is financed by an undisclosed investment from Australia's Medical Research Commercialisation Fund and private funding raised by the founders and directors.

In addition to Darling, Longas counts UTS professor Catherine Burke and Quadram Institute Director Ian Charles as cofounders. It was their research, and a frustration of not being able to do what they wanted to do, that drove the development of Morphoseq, which started in 2015, McCooke said.

Darling explained that he and Burke had been collaborating on better ways of profiling microbial communities using the 16s rRNA gene, which contains a 1.5-kb-long stretch of repeats that couldn't be resolved using short-read sequencing.

"That's when we developed this first prototype synthetic long read tech," Darling said. "We came to the realization this could be useful if applied elsewhere."

Darling said Longas paid close attention to the problem of repeats when designing Morphoseq. "We are able to scan all major classes of repeats in bacteria and human, which can be 6 to 7 kilobases. We chose [10-kb-long reads] because it gets us through all the major classes of repeats in the organisms," he said.

In Goldman's 2012 paper, his team suggested that around 10 kb was the maximum length that could be handled in an actual NGS analysis-by-mutatgenesis experiment.

The key principle behind Morphoseq is the introduction of a unique identifier to every molecule, accomplished through random mutagenesis and driven by mutagenic nucleotide analogs in strand synthesis and limited cycle PCR. "What's special and allows it to work is that we're able to introduce these mutations with a high degree of uniformity," Darling said.

"Mutations at all four nucleotides occur with basically equal probabilities," he added. "That's important because it maximizes the amount of information in the barcodes. We tend not to call them barcodes. That tends to push people into thinking they're synthetic, we call them molecular identifiers. They're created in the process, rather than introduced as an external synthetic barcode."

With Morphoseq, "any repeats [in the template] get mutated out," Darling said, and the mutation pattern enables algorithmic linking to recreate the longer template. "Conceptually it's simple, but implementing it is

challenging. There are tricky chemistry problems that needed to be solved and a whole bunch of work to be done in algorithmics."

On the other end of the sequencing pipeline from the library prep kit lies Longas' proprietary cloud-based software. Darling said it can assemble genomes as well as map reads.

"The assembly is part of the pipeline," Darling said. Everything can be automated through to the assembly, but the software also can provide virtual long reads or short reads. "It's quite easy for researchers to have introspection into the data and convince themselves something hasn't gone wrong along the way," he said.

When asked if Longas would make the software open source, McCooke punted. "We haven't really determined that," he said. "As it stands, there are proprietary elements to it. It's a question of market acceptance. I think our approach is to say there's a proprietary component to this, but we'll test that with the market."

So far, Longas' validation work has consisted of testing the kit with various sequencing platforms. The firm said it has gotten it to work with Illumina and BGI sequencers. The design intention is that it would be applicable to any short-read sequencer," Darling said. "But we haven't validated them all yet," due to resources. He said the firm is also looking to validate the kit on Thermo Fisher Scientific's IonTorrent instruments, and has not yet considered Qiagen's GeneReader, but suggested there was no reason it couldn't work.

Darling added that Longas designed Morphoseq so that it could be used in any NGS lab. "One design goal was simplicity," he said. "It fits the model of a small reagent kit. There's nothing special or unusual about it."

To get the kit into NGS labs, McCooke said he was exploring deals with other companies. "We would prefer to partner to get this into the market," he said. "We could make arrangements with the sequencing companies themselves. There are third parties as well that sell into the sequencing market."

McCooke said Longas has yet to determine pricing, but suggested that if the company were to take Morphoseq to market itself, the kit and software would be a single cost.

Despite being in stealth for years, McCooke said there was "no shortage of people interested" However, he declined to name any of those parties. He later added, "there has been most interest from research groups looking at microbial genomes, including the microbiome and environmental isolates, and cancer genomics." The company will also provide access to a "small number" of validation testers, he said.

When pressed for how it compared to competing technologies from 10x Genomics and BGI, Darling noted that Morphoseq "doesn't need to be sold with an instrument" and isn't tied to a particular platform.

In general, he claimed Morphoseq is scalable, cost effective, easy to use in the lab, and provides results that "are really good."

"About 92 percent of the reads are above Q30," Darling said. And the remaining 8 percent? "There's a tail that goes down pretty quickly to about Q20," he said. "There's nothing below Q20." He added, "some of this is subject to change and doesn't reflect the limits of the technology [...] there's still a lot of room for improvement."

Darling also suggested that internal testing shows Morphoseq could help sequence degraded DNA samples that might not be suitable for a long-read sequencer.

Darling said he would feel Longas were successful when he sees sequence databases "filling up with completed genomes rather than contigs."

"If you look at what's piling up in databases, there are not many long-read sequences or finished genomes," he said. "The vast majority are from Illumina."

Filed Under

[Sequencing](#)

[Informatics](#)

[Sample Prep](#)

[Business News](#)

[product launch](#)

[NGS](#)

Subscribe

[Privacy Policy](#). [Terms & Conditions](#). Copyright © 2019 GenomeWeb LLC. All Rights Reserved.