

## Promise and Pitfalls of Third-Generation Sequencing

07/25/2012

Janelle Weaver, Ph.D.

*Why hasn't third-generation sequencing technologies taken off just yet? Janelle Weaver reports on the uphill battle lying ahead for these systems.*

Like

Tweet

0

0

1

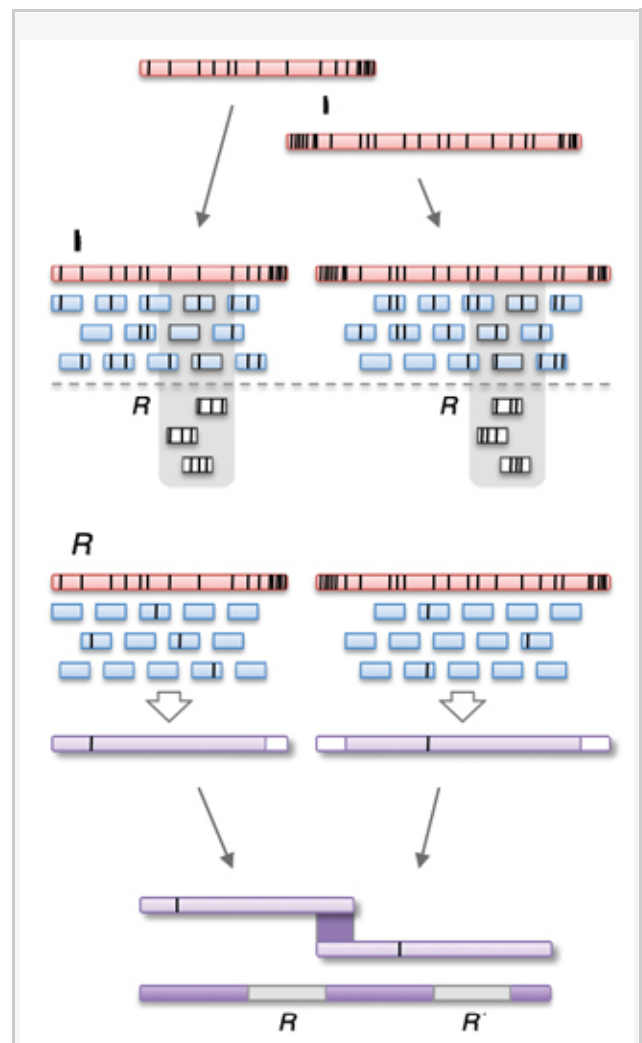
When Pacific Biosciences announced last April that it would start shipping its commercial PacBio RS systems, the company expected that the release of this third-generation sequencing product would "immediately expand the applications of DNA sequencing in fields such as cancer research, pathogen detection, and agriculture." Unlike second-generation systems on the market, the PacBio RS system was capable of carrying out single-molecule sequencing reactions in real time, producing results within a day. Moreover, the long sequence reads—spanning several thousand DNA bases—would enable *de novo* sequencing, simplify sequence assembly by spanning repeat regions, and enhance the detection of copy number variations. Because no DNA amplification is required, the system would reduce certain artifacts and biases in genome coverage.

But now, over a year later, it seems that researchers have been slow to adopt third-generation sequencing technology. Why? Well, one disadvantage had been relatively high error rates. Although high accuracy can be achieved through circular consensus sequencing (CCS), which involves sequencing shorter templates multiple times, the PacBio RS instrument generates single-pass reads that average less than 85% nucleotide accuracy. "The high error rate of these sequencing reads has been seen as the primary limitation of the technology," says Adam Phillippy, a genomics investigator at the National Biodefense Analysis and Countermeasures Center.

To circumvent this problem, Phillippy and his collaborators developed a new hybrid method that combines second- and third-generation sequencing approaches to yield almost perfect accuracy for long reads, as reported July 1 in *Nature Biotechnology* (1). Applied to the parrot genome among others, their method corrected individual long-read sequences by first mapping short-read sequences to them and computing a highly accurate hybrid consensus sequence. Short reads were produced by 454 and Illumina instruments, as well as PacBio RS CCS, while long single-pass reads were produced by PacBio RS. "We developed the first algorithm capable of correcting and assembling PacBio RS single-molecule sequencing reads and demonstrated that the high error rate of the PacBio RS technology can be managed to greatly improve genome and transcriptome assembly," says Phillippy.

### Peaceful Coexistence?

But there's still a lot of work to be done. For example, software developers need more time to catch up to the new instruments. "Third-generation instruments are generating an entirely new type of sequencing data," says



One reason that researchers have not been quick to adopt third-generation sequencing technology is its high error rate. Phillippy and his collaborators developed a new hybrid method that combines second- and third-

Phillippy. "The past five or more years of algorithm development have focused almost entirely on high-throughput, high-accuracy, short-read data. It takes considerable time to turn the software development process around to a new focus." Phillippy's algorithm is a step in the right direction because the corrected reads can now be analyzed by existing bioinformatics tools that couldn't handle the high error rate.

The technology also needs to improve its reliability, throughput, and costs before it can become competitive, says Phillippy. "There were similar lag times of two to three years between the introduction of 454 and Illumina technologies before they became widely accepted and pushed Sanger sequencing to a niche role."

Pacific Biosciences is in the process of improving the instrument's throughput and extending read lengths, says Edwin Hauw, the company's senior director of product management. "The system hardware itself is not changing, but we're improving the chemistry and software," he says. Currently, the system is well suited for studying microbial genomes, but its throughput limits the study of larger genomes. "It is cost prohibitive for certain applications, so targeted sequencing for human genomes or other large genomes is the best strategy for the time being," says Hauw.

Once these hurdles are overcome, the new technology will enable researchers to gain a deeper understanding of many diseases—such as cancer, autism, and chromosomal disorders—that are associated with copy number variations and other large-scale structural variations that can't be easily probed with second-generation sequencing technology. Long single-molecule sequences can also reveal insights into "junk DNA" contained in non-coding intronic and intergenic regions of the genome, which are thought to play important regulatory functions but have not yet been studied extensively because they could not be properly assembled.

But third-generation sequencing technology is unlikely to replace prior technology anytime soon. In the end, the choice of sequencing technology will depend on the specific research question. For example, population studies that require a high depth of sequencing, such as human single-nucleotide polymorphism surveys or expression studies, will remain best studied with second-generation technologies that produce huge amounts of data at a very low cost. "Until the third-generation technologies can match this cost-per-base, they will be limited to applications where read length is particularly important, like genome assembly or structural variation studies," says Phillippy. "I expect that second- and third-generation technologies will peacefully coexist until there is another sea change."

## Far from Mature

generation sequencing approaches to yield almost perfect accuracy for long reads. Source: Nature BioTechnology

### Recent Activity

**Sign Up**

Create an account or **log in** to see what your friends are doing.



**BioTechniques - Ph.D. Students Losing Interest**

36 people recommend this.



**BioTechniques - Calendar Girls**

19 people recommend this.



**BioTechniques - Future Methods in PCR**

13 people recommend this.



**BioTechniques - Calendar Girls**

19 people recommend this.

Facebook social plugin

Like

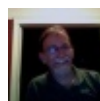
Send

Sign Up to see what your friends like.



Add a comment...

Comment using...



**Charles M. Nicolet** · Bowdoin College

"Sanger technology use has been decreasing dramatically, and the end may be near for Sanger sequencing" love these predictions! Go talk to someone who works in a Sanger core and ask about that.

Reply · Like · 2 hours ago

Facebook social plugin



BioTechniques **MyBioTechniques**

**MyBioTechniques** WIRED launches health conference.

Keynote speaker: Craig Venter. Read more:

[awe.sm/c1rSq](http://awe.sm/c1rSq)

about 1 hour ago · reply · retweet · favorite

**MyBioTechniques** What if we could model drug toxicity

and effectiveness on a chip? We will, soon. Learn

more: [eepurl.com/nYj5n](http://eepurl.com/nYj5n)

about 1 hour ago · reply · retweet · favorite

**MyBioTechniques** Why hasn't third-generation

sequencers taken off yet? Find out here:

[eepurl.com/nX6vb](http://eepurl.com/nX6vb)

2 hours ago · reply · retweet · favorite

**MyBioTechniques** FASEB: House FY2013 funding bill

will delay research progress at the NIH. Read the

As director of technology development at the Advanced Technology Program at the National Cancer Institute-Frederick and SAIC-Frederick, David Munroe is charged with acquiring promising new sequencing technology. Having recently acquired a PacBio RS system, which just went into service, he recognizes the advantages of the system, including fast turnaround times. Munroe says it's just a matter of time before people switch over. "It takes a while for people to get used to the type of data that comes off the machine and get the software packages in place for analyzing that data."

But from a technological standpoint, third-generation systems that are commercially available, such as those offered by Helicos BioSciences and Pacific Biosciences, are far from mature, according to Fatih Ozsolak, a scientist at Helicos BioSciences. "It is clear that they provide several unique capabilities that cannot be matched by second-generation sequencing technologies, but overall, they cannot meet the demands of today's research and diagnostics market as well as second-generation sequencing technologies do," she says. These companies entered the market at an early stage because of economic pressures, not because the new technology was superior to existing technology (2).

From a business standpoint, third-generation sequencing technology companies are at a disadvantage. "With significant investments from pharmaceutical and academic research facilities, particularly during the past three years, the sequencing market may already be saturated with high-throughput second-generation sequencing technology sequencers," says Ozsolak. "Even major second-generation sequencing technology providers expect much of their sales in the near future to be driven by the lower-cost benchtop second-generation sequencing technology systems that are more suited to budget-conscious research environments and clinical diagnostic facilities."

But in 10-15 years, Ozsolak believes third-generation systems will eventually replace second-generation ones as the technology matures and more companies enter the market. "I am sure that 15-20 years ago, there were some who considered Sanger technology as a big leap and thought there might always be a place for it, but as the second-generation sequencing technologies have become more and more affordable, Sanger technology use has been decreasing dramatically, and the end may be near for Sanger sequencing," she says. "Third-generation sequencing technologies are in a position where second-generation sequencing technologies were in the early 2000s."

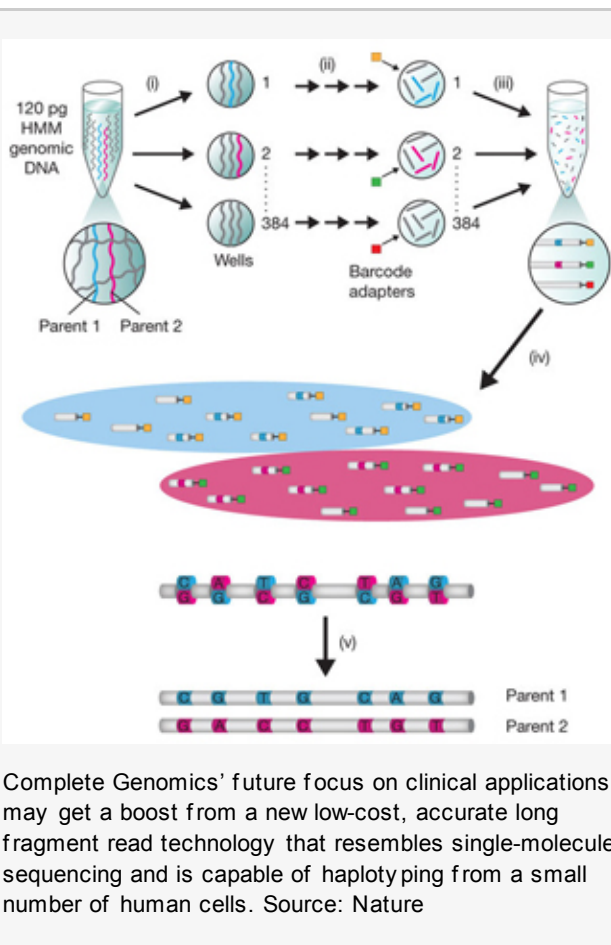
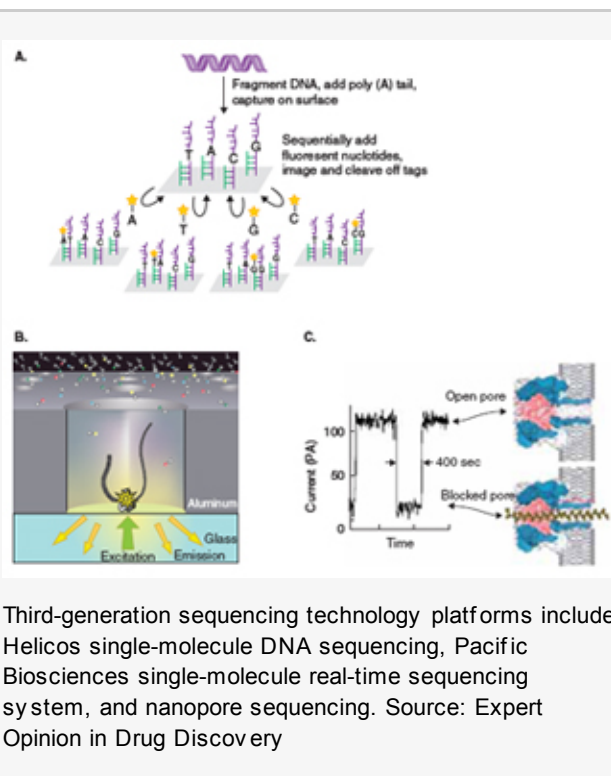
### Clinical Impact

Clifford Reid, co-founder and CEO of Complete Genomics, also believes that single-molecule sequencing technologies will not replace prior technology anytime

press release: [awe.sm/d1rxj](http://awe.sm/d1rxj)  
3 hours ago · reply · retweet · favorite



Join the conversation



soon. "We, as a scientific community, don't yet have all of the tools and techniques that we need to effectively handle single molecules. There's still a long way to go," he says.

An expert in software development, Reid launched Complete Genomics in 2008 after crossing paths with a biologist named Radoje Dmanac, who became co-founder and chief scientific officer of the company. They combined their distinct areas of expertise to form a company that provides whole human genome sequencing and analysis services to researchers. Now the company is focusing on getting certification to offer clinical services.

The company's future focus on clinical applications may get a boost from a new method Dmanac and his collaborators reported July 12 in *Nature* (3). The researchers developed a low-cost, accurate long fragment read technology that resembles single-molecule sequencing and is capable of haplotyping from a small number of human cells. The data allow haplotypes to be assigned to maternal and paternal lineages, and this information is important for incorporating parental imprinting in genetic diagnoses.

"This is a huge step forward. The long read lengths will allow us to do a lot of things that we couldn't possibly do before, and the reduced amount of input DNA is helpful," says Munroe. Input DNA requirements for sequencing can be fairly high, which has been a hurdle for clinical applications.

Tailoring sequencing technologies, whether second- or third-generation, to the clinical arena could have a profound impact on medicine. "The genome is different in a really fundamental way than every other diagnostic test, because it doesn't change over your lifetime, whereas other tests take a snapshot of you at a moment in time," says Reid. Until now, rapid improvements in sequencing technology has continually made existing genomic data obsolete, so scientists have simply resequenced genomes with newer technology rather than rely on old data gathered with prior technology. But moving forward, genetic test results could be permanently stored in clinical databases. "The genetic diagnostics world is right at the beginning of being turned completely upside down, and I think it's going to take a lot of time. It's not going to happen overnight, but the direction is pretty inevitable just because of this fundamental quality of the human genome—that it's finite, and we are going to get it right."

## References

1. Koren, S., M. C. Schatz, B. P. Walenz, J. Martin, J. T. Howard, G. Ganapathy, Z. Wang, D. A. Rasko, W. R. McCombie, E. D. Jarvis, and A. M. Phillippy. 2012. Hybrid error correction and *de novo* assembly of single-molecule sequencing reads. *Nature Biotechnology* doi: 10.1038/nbt.2280.
2. Ozsolak, F. 2012. Third-generation sequencing techniques and applications to drug discovery. *Expert Opinion on Drug Discovery*. 7(3):231-243.
3. Peters, B. A., B. G. Kermani, A. B. Sparks, O. Alferov, P. Hong, A. Alexeev, Y. Jiang, F. Dahl, Y. T. Tang, J. Haas, K. Robasky, A. W. Zaranek, J. H. Lee, et al. 2012. Accurate whole-genome sequencing and haplotyping from 10 to 20 human cells. *Nature* 487(7406):190-195.

Keywords: sequencing



[privacy](#) [terms & conditions](#) [feedback](#) [contact us](#)

[Close Window](#)