

# Born-Again Sequencing

By Aaron J. Sender

**B**y resurrecting a genome analysis technique that hadn't been performed in nearly a decade, Daniel Peterson promises to drastically slash the cost of genome sequencing. "The savings would be in the hundreds of millions of dollars — maybe even more," he says.

The technique, called Cot analysis, was the method of choice for extracting information such as genomic size and sequence complexity in the 1970s, but was left for dead with the advent of molecular cloning methods. "For a long time it was the only way to study genomes," says Peterson. "In fact, most of what we know today about genome structures comes from these Cot analysis studies."

Cot analysis uses kinetics to measure the relative occurrence of specific sequences in the genome. Genomic DNA fragments are denatured into single strands and then allowed to renature. The idea is that the more copies of a particular sequence in a genome, the faster it will find a partner and renature. In this way researchers can isolate the DNA fragments into three distinct bins: highly repetitive, moderately repetitive, and single/low-copy DNA.

"If you have a genome that's 80 percent repetitive DNA, and you do shotgun sequencing by trying to pull clones out at random and sequencing them, 80 percent of the time you're just going to be sequencing the same repeats over and over again," says Peterson.

Cot-based cloning and sequencing

(CBCS), however, eliminates the randomness associated with shotgun sequencing, sharply reducing the number of clones needed for complete coverage of all unique sequences. For example, the onion genome would require 119 million clones for shotgun sequencing. For CBCS, on the other hand, a mere 16 million would do. Assuming a cost of \$3.40 to prepare and sequence each clone, that's a saving of more than \$350 million.

Today when Peterson describes his approach, a common response is, "It's just so obvious. Why didn't people think of this before?" he says. But when he first performed a Cot analysis on the tomato genome as a Colorado State grad student in 1998, genome sequencing didn't enter his mind. He simply wanted to isolate the low-copy DNA, presumably genes, and use them as probes to find their locations on the chromosomes.

When he submitted a paper on the work, it was turned down. "One of the reviewers said, 'This is an antiquated technique. Why would anybody do this?'" says Peterson.

It wasn't until he moved to Andrew Paterson's plant genome mapping lab at the University of Georgia as a postdoc that the potential of Cot analysis for genome sequencing began to emerge.



Daniel Peterson

But running a Cot analysis is not easy. "It's a very hard thing to do because you have to adhere to a certain series of very inflexible rules for your data to make any sense," says Peterson. For example, "you have to get extremely pure DNA. If it has any kind of contaminants in it at all, it can really screw things up," he says. Conventional DNA purification leaves too many proteins associated with the DNA. Interpreting Cot curves also requires complex mathematical calculations.

So Peterson, who is set to take on a faculty position at Mississippi State next month, is toying with the idea of starting a company that would offer CBCS as a service.


"Sequencing all that repetitive DNA is just not economically feasible," he says. "There are various groups that want this genome or that genome sequenced. [CBCS] would make genome sequencing available for a lot of species that would otherwise not be possible." 

Photo: Michael Paul Elfand