Improving the accuracy of long-read genome sequencing

13 August 2019, by Bob Yirka

Credit: CC0 Public Domain

A team of researchers from institutions in the U.S., Germany and China has developed a way to improve the accuracy of long-read genome sequencing. In their paper published in the journal *Nature Biotechnology*, the group outlines how they improved an existing technique, and how well it works.

The researchers began their work by noting that the current status quo in DNA sequencing technology allows for sequencing short sequences (short reads) accurately, or long sequences (long reads) less accurately. In this new effort, they sought to improve the accuracy of long-read sequencing. More specifically, they sought to improve the accuracy of a circular consensus sequencing (CCS) technique using single-molecule, real-time (SMRT) sequencing as performed by Pacific Bioscience (PacBio). Most of the team working on the project were from PacBio—other members were from Google, Stanford University, the Max Planck Institute, Saarland University, DNAnexius, NIST, the National Human Genome Research Institute, the Chinese Academy of Agricultural Sciences, the Dana-Farber Cancer Institute and Johns Hopkins University.

The CCS technique (which is not considered a long-read technique) currently in use by PacBio involves the use of hairpin adapters, which are ligated onto the ends of DNA molecules to create a template. A polymerase starts at the adapter and moves across the DNA piece, adding bases as it goes, creating the read. The technique also involves moving the polymerase across the DNA bit multiple times. The researchers note that the CCS technique has typically been used with base lengths of 1000 to 2000 bases. By improving the technique, the team was able to read up to 10,000 bases accurately.

The researchers report that their improvements were based mostly on improving the quality of DNA used at the start of sequencing. The quality of DNA was improved by starting the reactions before loading the DNA and then waiting for up to an hour—if they found the polymerase was still moving, it had to be of higher quality. They also made sure that the DNA molecules were all of uniform size—they did this by using a SageELF instrument.

The team reports that the improved technique generated long reads (averaging 13.5 kilobases) with a 99.9 percent accuracy rate.

More information: Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome, *Nature Biotechnology* (2019). DOI: 10.1038/s41587-019-0217-9

© 2019 Science X Network

APA citation: Improving the accuracy of long-read genome sequencing (2019, August 13) retrieved 19 September 2020 from https://phys.org/news/2019-08-accuracy-long-read-genome-sequencing.html