New cost-effective genome assembly process developed

May 5 2013

DOE JGI researchers are part of a team that has developed what is described as "a fully automated process from DNA sample preparation to the determination of the finished genome." Credit: Roy Kaltschmidt, LBNL

The U.S. Department of Energy Joint Genome Institute (DOE JGI) is among the world leaders in sequencing the genomes of microbes, focusing on their potential applications in the fields of bioenergy and

environment. As a national user facility, the DOE JGI is also focused on developing tools that more cost-effectively enable the assembly and analysis of the sequence that it, as well as other genome centers, generates.

Despite tremendous advances in cost reduction and throughput of DNA sequencing, significant challenges remain in the process of efficiently reconstructing genomes. Existing technologies are good at cranking out short fragments (reads) of DNA letters that are computationally stitched back together (assembled) into longer pieces, so that the order of those letters can be determined and the function of the target sequence discerned. However, genome assembly, the equivalent of trying to put together a multi-million piece jigsaw puzzle without knowing what the picture on the cover of the box is, remains challenging due to the very large number of very small pieces, which must be assembled using current approaches.

As reported May 5 online in the journal *Nature Methods*, a collaboration between the DOE JGI, Pacific Biosciences (PacBio) and the University of Washington has resulted in an improved workflow for genome assembly that the team describes as "a fully automated process from DNA sample preparation to the determination of the finished genome."

The technique, known as HGAP (Hierarchical Genome Assembly Process), uses PacBio's single molecule, real-time DNA sequencing platform, which generates reads that can be up to tens of thousands of nucleotides long, even longer than those provided by the workhorse technology of the Human Genome Project era, the Sanger sequencing technology, which produced reads of about 700 nucleotides. The Sanger process involved creating multiple DNA libraries, conducting multiple runs, and combining the data, so that gaps in the code were covered and accuracies of a DNA base assignment were very high. Post-Sanger methods still typically require multiple libraries and often a mix of

technologies to produce optimal results. Instead, with HGAP, "only a single, long-insert shotgun DNA library is prepared and subjected to automated continuous long-read SMRT sequencing, and the assembly is performed without the need for circular consensus sequencing," the team reported.

This de novo assembly method was tested using three microbes previously sequenced by the DOE JGI. The data collected were compared against the reference sequences for these microbes and the team found that the HGAP method produced final assemblies with >99.999% accuracy.

"We are always on the lookout for new approaches that will improve upon the efficient delivery of high-quality data to our growing community of researchers," said Len Pennacchio, DOE JGI's Deputy Director of Genomic Technologies. "This technique is one of many improvements that we are pursuing in parallel to achieve additional economies of scale."

The DOE JGI's sequencing efforts account for more than 20% of the more than 20,000 worldwide genome projects (microbes, plants, fungi, algae, and communities of microbes) completed or currently in the queue, and most of those are focused on the biology of environmental, energy, and carbon processing.

"We enjoyed a very productive collaboration with JGI on this project and benefitted tremendously from the expertise of JGI's scientists in both the fields of microbiology and microbial genome assembly and annotation," said Jonas Korlach, Chief Scientific Officer at Pacific Biosciences. "This expertise provided us with the ability to adapt our single molecule sequencing assembly methods to produce a higher level of finished quality than was previously possible using a gold-standard Sanger finishing approach, and at a speed and price point competitive

with alternative next generation sequencing and assembly methods. We look forward to seeing what scientific advances will be enabled by this method as JGI's User Community assesses JGI's capabilities to assemble their microbial genomes using this new approach."

The team will now seek to extend the utility of this new assembly method beyond microbes to the genomes of more complex organisms.

More information: dx.doi.org/10.1038/nmeth.2474

Provided by DOE/Joint Genome Institute