Combining 454 Sequencing and Traditional Sanger Reads for Microbial Genomes

Feng Chen, Joseph Alessi, Edward Kirton, Jamie Jett, Eugene Goltsman, Alla Lapidus, and Paul Richardson

Genomic Technologies
DOE Joint Genome Institute
2800 Mitchell Dr. B400
Walnut Creek, CA 94598
fchen@lbl.gov
(925) 296-5733

The US DOE Joint Genome Institute (JGI) is a high-throughput sequencing center involved in a myriad of sequencing projects. A major effort at JGI is the sequencing of microbial genomes of relevance to the DOE missions of carbon sequestration, bioremediation and energy production. The JGI Microbial Program is responsible for the generation of over 200 microbial genomes and we are interested in utilizing new technologies to increase capacity. The 454 sequencing platform is an integrated system of emulsion-based PCR amplification of hundreds of thousands of DNA fragments linked to high throughput parallel pyrosequencing in picoliter-sized wells. The 454 sequencing platform can deliver 30 to 50 million base pairs (mbp) from a single run, however, our previous study revealed that the quality of the resulting assembly contains high numbers of misassemblies and base errors due to short read length and lack of paired-end information. The traditional Sanger sequencing method is lower in throughput and more costly but it provides high quality sequencing results and more accurate assemblies. The paired-end information from Sanger sequencing is proven to be crucial in scaffolding and gap closure. We combined 454 sequencing results with different amounts of paired Sanger sequencing reads from three different sized shotgun libraries and analyzed the results. Assemblies from all combinations were done by Newbler and Phred/Phrap and viewed and analyzed by Consed and in-house developed software. Numbers of remaining gaps and low quality regions from different combinations were assessed. Distribution of coverage and possible errors were analyzed for both platforms. We will also discuss the optimal ratio of data from 454 and Sanger sequencing to achieve high quality finished microbial genome sequences in a time and cost effective manner.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36.