LBNL-62592 Abs
UCRL-ABS-230191

Assessment of 454 Sequencing Errors in Microbial Genomes

Stephan Trong[1], Patrick Chain[2], Ed Kirton[1], Eugene Goltsman[1], Brian Foster[1], Paul Richardson[1] and Alla Lapidus[1]

[1]U.S. Department of Energy Joint Genome Institute, Walnut Creek, CA
[2]Lawrence Livermore National Laboratory, Livermore, CA

The introduction of pyrosequencing-based sequencing platforms such as 454-sequencing along with the Sanger technology for whole-genome shotgun sequencing has provided an alternative way to produce cost-effective sequence data. For microbial genome finishing, the 454/Sanger hybrid approach has made a considerable impact in reducing the time required to close a genome. However, errors within the 454 sequencing data remain to be addressed to make sure that final consensus represents high quality sequence.

JGI is working on further cost reduction of sequencing and finishing processes and is planning significant cut of Sanger sequencing. This will increase the portion of the genome covered with pyrosequence only. Thus the analysis of errors produced by this new technology becomes very important.

We conducted a study to assess the quality of the 454 sequence data in order to determine its impact on error rate, and to devise a strategy to correct or reduce the errors. The study analyzed 29 microbial genomes containing both Sanger and 454 only assemblies to look for mismatches in the 454 sequence data. We examined both homopolymer tracts and non-homopolymer regions with respect to read depth and quality assignment. An overview of our findings and strategy for reducing the 454 errors will be presented.

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. DE-AC52-06NA25396.