# "Cloneview": An assembly viewer for microbial genome Finishing

S.Trong[1], E. Goltsman[1], S. Malfatti[2], P. Chain[2], and A. L. Lapidus[1]

[1] Joint Genome Institute Production Genomics Facility, Walnut Creek, CA
[2] Lawrence Livermore National Laboratory, Livermore, CA

The process of improving and completing (Finishing) draft genome assemblies using whole genome shotgun approach involves many complex iterative steps. The effectiveness of this process depends on many factors, such as the complexity of the genome, the quality of the DNA sequence and the accuracy of the assembled data. Although many efforts have been attempted to automate Finishing, it is still a relatively manual process. One such effort involves the identification and correction of regions that are incorrectly pieced together by the assembly program. To aid the finishers in visualizing and identifying these problematic areas, we have developed a visualization tool called Cloneview for displaying mate pair information in an assembly.

Our goal was to provide a graphical tool for displaying assembly information in a way that could be used to easily identify misassembled regions, repeat regions and possible biases in library creation.

Cloneview's ability to highlight read pairs that are inconsistently placed in the assembly allows the finishers to easily detect regions that are misassembled. By extracting the names of the reads to a file, further processing to correct the misjoined areas can be achieved with minimal effort. Other features of Cloneview include displaying read and clone depth, read pairs that span gaps, library-specific reads and repeat regions. In addition to the viewer, we have developed an accompanying software program to detect and report misassembled regions by looking for areas where violations in mate pairing predominately outweigh valid ones. This tool along with the viewer further enhances the finishers' ability to resolve misassemblies more efficiently.

The Microbial Finishing groups at the JGI /PGF and JGI/LLNL have incorporated these tools into their pipeline as part of an effort to rapidly finish microbial genomes.

This work was performed under the auspices of the US DOE of Science, Biological and Environmental Research Program, and by the University of California, LLNL under Contract No. W-7405-Eng-48, LBNL under Contract No. DE-AC02-05CH11231 and LANL under Contract No. W-7405-ENG-36.