

Lawrence Berkeley National Laboratory

Recent Work

Title

Quick Draft Assembly Improvement for Improved Finishing

Permalink

<https://escholarship.org/uc/item/05k0k4h8>

Author

Copeland, Alex

Publication Date

2006-03-27

Quick Draft Assembly Improvement for Improved Finishing

Alex Copeland

Production Genomics Facility, Joint Genome Institute, Walnut Creek, CA

While prevention is a more cost effective means toward quality improvement than correction, in production sequencing, working with contaminated DNA samples is inevitable. Although we cannot always control the quality of the DNA we agree to sequence, we can control, to some extent, contamination in assemblies we pass along to finishing. We have noted that relatively modest improvements to assembler input sometimes makes large improvements in the resulting assemblies which has direct benefits for all users of the data, especially finishers.

As part of an effort to improve the quality of sequence data released by the Production Genomics Facility we created a process, and supporting software we call 'QD', to quickly identify and exclude low quality and contaminant reads from assemblies. We have applied this process to all historical microbial projects, and it is used on all new projects before they are sent for finishing.

The initial development of this process was driven by recognition that the JGI had draft sequenced a large number of microbes, whose data was not being utilized to its full potential by the research community, at least in part, due to the low quality of the assemblies. During development, we focused on the twin goals of keeping the process sufficiently simple as to make it possible to apply to all of our legacy projects, and doing no harm. The end result has made a significant difference in the quality of our assemblies and has simplified finishing these genomes.

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.

LBNL-59930 Abs.