

Abstract:

An Annotation ‘Minipipe’ to Rapidly Assess Genomic Assemblies of 454 pyrosequencing Reads

Alan Kuo¹, Darren Platt², Paul Richardson¹, Igor Grigoriev¹

The JGI Annotation Pipeline is routinely used to predict genes in assemblies of Sanger-sequenced eukaryotic genomes. In this study we instead use annotation to assess competing assemblies of a single genome sequenced using alternative technologies. These technologies promise to vastly lower the cost and thus expand the output of nucleotide sequencing compared to the traditional Sanger method. However, the new technologies also bring new challenges such as short reads and new kinds of sequence errors. At JGI we are exploring ways to incorporate 454 pyrosequencing into the standard JGI genome workflow. One way to exploit the respective strengths of 454 and Sanger sequence may be to derive ‘hybrid’ assemblies from both. To rapidly and consistently assess hybrid, 454-only (20X), and Sanger-only (4X) assemblies of the genome of the ubiquitous plant pathogen *Phytophthora capsici*, we developed a short version of our standard JGI Annotation Pipeline. This ‘minipipe’ predicts genes, and then examines 1) internal stop codons and 2) truncated homologs of genes of Sanger-only (8X) *Phytophthora* sp. Both metrics serve as surrogates for frameshifts resulting from sequencing errors or polymorphism. Our results suggest that while the rate of 454-related frameshifting is high, the hybrid assembly approach holds promise as a low-cost way of estimating the number of genes in a genome. We conclude that annotation is a quick and simple way to assess the quality of novel assemblies.

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 Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396

¹ Lawrence Berkeley National Laboratory

² Lawrence Livermore National Laboratory