



Category: Bioinformatics

Comparative assembly and analysis of different sized genomes using Pacbio sequencing technology

Ridhi Goel, Pooja Raj, C.P. Rajadurai and Dhinoth Kumar

Agrigenome, Kakkanad, Cochin, INDIA

Presenting author: ridhi.g@aggenome.com

Abstract

PacBio is the third generation sequencing technology which is based on the single molecule real time sequencing (SMRT) platform using the property of zero-mode waveguide (ZMW). This technology generates very long reads which is best suited for various applications like de novo genome assembly, structural variations, full length transcriptomes, direct detection of base modifications etc. PacBio data can either be used alone or in combination with the illumina based shorter reads to facilitate a good assembly. Different algorithms are available to construct the genome based on PacBio alone or hybrid datasets. In order to identify the best possible approach we did a comparative study employing the widely accepted assembly tools on *E.coli*, *C.elegans* and *A.thaliana* datasets (PacBio & Illumina (Paired end & Mate Pair)). We performed *de novo* genome assembly, gene prediction and gene annotation for all possible dataset (PacBio & Illumina PE & MP) and tools combination. The study resulted in the identification of the best method that could assemble the 4.6 MB of *E.coli* genome covering ~97% of BUSCO represented genes in a single contig. For *C.elegans* and *A.thaliana* we were able to achieve 109 MB and 123 MB sized assembly with ~80% of BUSCO represented genes.

Citation: Goel, R., Raj, P., Rajadurai, C.P. and Kumar, D. Comparative assembly and analysis of different sized genomes using Pacbio sequencing technology [Abstract]. In: Abstracts of the NGBT conference; Oct 02-04, 2017; Bhubaneswar, Odisha, India: Can J biotech, Volume 1, Special Issue (Supplement), Page 231. <https://doi.org/10.24870/cjb.2017-a216>