Hybrid de novogenome assembly using Illumina, PacBio and Hi-C sequencing data

Hansheng Zhao

Abstract

Citation: Hansheng Zhao Hybrid de novogenome assembly using Illumina, PacBio and Hi-C sequencing data. **protocols.io**
dx.doi.org/10.17504/protocols.io.rf4d3qw
Published: 03 Jul 2018

Protocol

Data preprocessing for Illumina data

Step 1.
Low-quality reads (the proportion of the base of Q<13 more than 40% in a given reads) were filtered using NGS QC Toolkit (version 2.3.3) with default parameters.

SOFTWARE PACKAGE (LINUX -)

NGS QC Toolkit, 2.3.3

Data preprocessing for PacBio data

Step 2.
For PacBio data, we used MECAT (release 20170627) to correct errors with the following parameters: -x 0 -i 0 -t 60 -r 0.8 -a 1000 -c 5 -l 2000.

SOFTWARE PACKAGE (LINUX -)

MECAT, released 2017062

Data preprocessing for HiC data

Step 3.
Valid HiC data were were evaluated and qualified using HiC-Pro (version 2.8.0_devel).

SOFTWARE PACKAGE (LINUX -)

HiC-Pro, 2.8.0_devel

Platanus

Step 4.
Platanus (version 1.2.4), an *de novo* and high heterozygous genome assembler, was carried out to assemble the fragment PE reads into contigs by constructing De Bruijn Graphs with automatically optimized k-mer size.

SOFTWARE PACKAGE (LINUX -)

Platanus, 1.2.4

DBG2OLC

Step 5.

Then, the corrected PacBio reads and the assembled contigs were thrown into DBG2OLC (release 20150611) to construct scaffolds with the parameters: DBG2OLC Contigs contig.fa LD 0 K 17 KmerCovTh 4 MinOverlap 25 AdaptiveTh 0.007 RemoveChimera 1 f scaffold.fa.

SOFTWARE PACKAGE (LINUX -)

DBG2OLC, release 20150611

A polish process

Step 6.

A polish process before the SSPACE process referred to consensus analysis of DBG2OLC, which were contributed to enhance the quality of the genome assembly and reduce errors in the SSPACE process.

SSPACE

Step 7.

The assemblies were elongated by SSPACE (version 3.0) using the MP reads and some gaps were filled using the Illumina and PacBio data by GapCloser (version 1.12) and PBJelly (release 20150824).

SOFTWARE PACKAGE (LINUX -)

SSPACE, 1

3D-DNA pipeline

Step 8.

The valid Hi-C data together with the above assembly were processed by 3D-DNA pipeline (version 170123) to produce chromosome-level scaffolds. Then, the contact maps were visualized by JuiceBox (version 1.5.2).

SOFTWARE PACKAGE (LINUX -)

JuiceBox, 1.5.2