Draft genome assembly using parasitic mite population NGS DNA sample from mites extracted from host wound environment

Ehtesham Mofiz

Abstract

Citation: Ehtesham Mofiz Draft genome assembly using parasitic mite population NGS DNA sample from mites extracted from host wound environment. **protocols.io**
https://www.protocols.io/view/Draft-genome-assembly-using-parasitic-mite-populat-exwbfpe
Published: 09 May 2016

Protocol

Step 1.
Sequence DNA sample in Illumina HiSeq generating paired-end data.

DATASET
Illumina HiSeq 2500 paired-end sequence data

Step 2.
Assess read qualities of read files (fastq files) using FASTQC.

SOFTWARE PACKAGE (Linux)
FASTQC

Step 3.
Trim fastq read files for adapter and quality (Q>=20).

SOFTWARE PACKAGE (Linux)
Trim Galore!

Step 4.
Download host genome (e.g. human hg19). Build Bowtie2 reference index of the host genome. Align adapter and quality trimmed reads to the host reference genome using Bowtie2 (using end-to-end mode and --very-fast parameters).

SOFTWARE PACKAGE (Linux)
Bowtie2
DATASET
host genome

Step 5.
Extract unmapped paired reads using samtools view with -f 12 flag (read unmapped, mate unmapped)

SOFTWARE PACKAGE (Linux)
SAMtools
cmd COMMAND (Linux)
`samtools view`

Step 6.
Assemble paired-end unmapped reads using Platanus assembler (default settings). From the

assemblies, filter out scaffolds that are less than 500 bp to get major scaffold assemblies.

SOFTWARE PACKAGE (Linux)

Platanus

Step 7.

Download Microbial RefSeq Database from The National Center for Biotechnology Information (NCBI). Align Platanus scaffolds to the database using BLASTN (e-value cutoff 10^{-20}; max_target_seqs 1). Filter for top hits and filter out scaffolds that have more than 80% of their length aligned to bacterial sequences. This gives you the draft assembly.

SOFTWARE PACKAGE (Linux)

BLASTN

DATASET

Microbial RefSeq Database from The National Center for Biotechno