protocols.io

Assembly Introduction

Frank Aylward and Daniel Mende

Abstract

Provides a short introduction of using the MEGAHIT, IDBA-UD, and SPAdes assemblers.

Guidelines

Assessment of Results

Stat	MEGAHIT	metaSPAdes	IDBA-UD
# contigs (>1kb)	18,394	13,266	16,256
Length in contigs (>1kb)	194.4 Mb	195.3 Mb	194.8 Mb
N50	48,090	70,745	54,716
# predicted genes	192,693	189,672	192,394
% reads recruited	95.12%	98.21%	97.83%
# misassemblies	386	436	853
bp in misassemblies	11.4 Mb	22.0 Mb	18.4 Mb
Metagenome fraction (%)	89.7%	89.7%	89.9%

Protocol

Step 1.

Move to directory containing assemblers.

cmd COMMAND

```
$ cd /home/c-debi/ecogeo/assembly
```

Step 2.

View assembler parameters for MEGAHIT v1.0.3, IDBA-UD v1.1.1, and SPAdes v3.7.1

cmd COMMAND

```
$megahit

$idba_ud

$spades.py
```

These commands will show parameters for each assembler.

Step 3.

Test Data Set:

Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communites. Shakya et al. (2013) Environmental Microbiology

Artifical metagenome combining genomic DNA of 62 organisms. Illumina sequencing 5.5 million paired reads.

Step 4.

Trimmomatic Quality Control:

cmd COMMAND

```
$ java -
jar trimmomatic-0.35.jar PE SRR606249_1.fastq SRR606249_2.fastq R1_pe R1_se R2_pe R2_se ILL
UMINACLIP:TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:10:28 MINLEN:50
```

This step has already been completed for you and the command does NOT need to be executed again.

Step 5.

Assemble with Megahit:

cmd COMMAND

```
$ megahit --preset meta-sensitive -1 SRR606249.trim_1.fastq -2 SRR606249.trim_2.fastq -
o SRR606249.megahit_asm
```

This step has already been completed for you and the command does NOT need to be executed again.

Step 6.

IDBA-UD: merge FASTQ files to interleaved FASTA files

File_R1: >Seq1   File_R2: >Seq1

File_merged: >Seq1.1

              >Seq2.1

cmd COMMAND
```
$ fq2fa --merge --
filter SRR606249.trim_1.fastq SRR606249.trim_2.fastq SRR606249.trim.merged.fasta
```
This step has already been completed for you and the command does NOT need to be executed again.

Step 7.

Peform assembly using IDBA-UD:

cmd COMMAND
```
$ idba_ud -r SRR606249.trim.merged.fasta -o SRR606249.idbaud_asm --num_threads 45
```
This step has already been completed for you and the command does NOT need to be executed again.

Step 8.

Perform assembly using MetaSPAdes:

cmd COMMAND
```
$ spades.py -o ./SRR606249.spades_asm --
meta -1 SRR606249.trim_1.fastq -2 SRR606249.trim_2.fastq --threads 60 --memory 600
```
This step has already been completed for you and the command does NOT need to be executed again.

Step 9.

Reference assessment: QUAST can perform comparisons against the reference genomes used to construct artifiial metagenome. Start with a baseline size of contiges (>1kb).

cmd COMMAND
```
$ seqmagick convert --min-length 1000 final.contigs.fa megahit_SRR606249.min1000.fasta
```

Step 10.

QUAST against 62 reference genomes:

cmd COMMAND
```
$ metaquast.py megahit_SRR606249.min1000.fasta -R ../Shakya_RefGenomes/
```
This step has already been completed for you and the command does NOT need to be executed again.