Manual dissection of the Schistosoma mansoni head and back end for transcriptomic analysis

Alise Ponsero[1]

[1]University of Arizona

Oct 16, 2018

Working

dx.doi.org/10.17504/protocols.io.uk5euy6

iMicrobe Metafunc course 2018

Alise Ponsero
University of Arizona

ABSTRACT

How to run Megahit version 1.1.3 (Li et al. 2015) through the iMicrobe plaform.

Megahit is a de-novo assembler for large and complex metagenomics NGS reads relying on succinct *de Bruijn* graph (SdBG) to achieve low memory assembly.

More informations about Megahit can be found here : https://github.com/voutcn/megahit

TAGS

metagenomics

assembly

EXTERNAL LINK

https://www.imicrobe.us/#/apps/59

PROTOCOL STATUS

Working

We use this protocol in our group and it is working

GUIDELINES

More informations and details about Megahit can be found there : https://github.com/voutcn/megahit

Several parameters are available to the user through the iMicrobe app.

- Choosing k

MEGAHIT uses multiple *k*-mer strategy. Minimum *k* (*--k-min*) , maximum *k* (*--k-max*) and the step for iteration (*--k-step*) can be set by options. Please not that *k* must be odd numbers while the step must be an even number.

- Filtering (Kmin+1)-mer

The parameter *--min-count* allows the user to discard (k_{min}+1)-mer with multiplicity lower than *d*. By default that parameter is set to 2. For high sequencing depth coverage (>40x), this value should be set to 3.

- mercy k-mer

This tool is specially designed for metagenomics assembly to recover low coverage sequence. For generic dataset >= 30x, MEGAHIT may generate better results with using the *--no-mercy* option.

- K-min-1 pass mode

This mode is more memory efficient for ultra low-depth datasets. It can be activated by using the option *--kmin-1pass*

BEFORE STARTING

- You need a working Cyverse account to connect to iMicrobe.To explore how to log into iMicrobe, read the dedicated protocol.

- Your dataset of interest should be metagenomic reads, in a fasta or fastq format.

- In iMicrobe, there is several ways to run an app on a dataset (from the cart, from your personnal datastore and form an URL). If you need more information on how to run an app, read the protocol associated. This app is currently not available to run from the cart.

De Novo assembly of paired-end illumina dataset

1 *Note :This protocol uses a samples from the HMP project (SRS143565). This sample is a right cubital fossa WGS dataset from a healthy female subject.*

Note : Currently this app cannot use a cart imput. Use a Cyverse datastore or URL imput.

In the 'tools' dropdown menu, select 'Apps'. You are presented the list of apps currently available on iMicrobe. Click on imicrobe-megahit-0.0.2u1.

In the page app, open the datastore and provide the **forward read** and **reverse read** file in the corresponding imput fields. Additionally, the user can tune the assembly parameter by passing a string of arguments in the "**Additional Command Line Arguments**" field.

- *Choosing k*

MEGAHIT uses multiple *k*-mer strategy. Minimum *k* (*--k-min*) , maximum *k* (*--k-max*) and the step for iteration (*--k-step*) can be set by options. Please not that *k* must be odd numbers while the step must be an even number.

- *Filtering (Kmin+1)-mer*

The parameter *--min-count* allows the user to discard (k_{min}+1)-mer with multiplicity lower than *d* . By default that parameter is set to 2. For high sequencing depth coverage (>40x), this value should be set to 3.

- *mercy k-mer*

This tool is specially designed for metagenomics assembly to recover low coverage sequence. For generic dataset >= 30x, MEGAHIT may generate better results with using the *--no-mercy* option.

- *K-min-1 pass mode*

This mode is more memory efficient for ultra low-depth datasets. It can be activated by using the option *--kmin-1pass*

After the job is effectively ran, you can access your results using the drop-down menu 'Tools' and selecting 'Jobs'. Select the job corresponding to your centrifuge run, and go to the section 'Outputs'. The centrifuge output files are now in your cyverse datastore. Click on 'Browse and view output files in the CyVerse Datastore'.

In the job folder created in the CyVerse datastore, the input fasta/fastq files are copied, along with the logs of the job (*.err and *.out). In order to retrieve your results go to the **megahit-out** folder. It contains the following outputs:

- **final.contigs.fa** : This fasta file contains the assembled contigs.
- **log** and **opts.txt** : Those text files contains the running log of the tool and the parameters chosen by the user respectively.
- the folder **intermediate_contigs**:

k*K*.contigs.fa contains the contigs assembled from the de Bruijn graph of order-***K***, they can be converted to a SPAdes-like FASTG file for vizualisation
k*K*.addi.fa contains the contigs assembled after iteratively removing local low coverage *unitigs* in the de Bruijn graph of order-***K***
k*K*.local.fa contains the locally assembled contigs for k=***K***
k*K*.final.contigs.fa contains the stand-alone contigs for k=***K***

EXPECTED RESULT

final.contigs.fa

De Novo assembly of 454 datasets

2 *Note : Currently this app cannot use a cart imput. Use a Cyverse datastore or URL imput.*

In the 'tools' dropdown menu, select 'Apps'. You are presented the list of apps currently available on iMicrobe. Click on imicrobe-megahit-0.0.2u1.

In the page app, open the datastore and provide the reads in the imput field "**single-end reads**".
Additionally, the user can tune the assembly parameter by passing a string of arguments in the "Additional Command Line Arguments" field. Those arguments are described in step 1.

The outputs from the app are similar to those described for paired-end datasets in step 1.