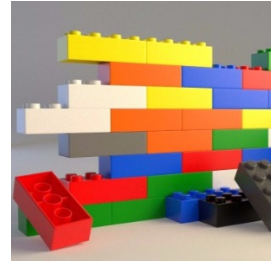


Oct 10, 2017 Version 2

Assembly with Megahit V.2

DOI

dx.doi.org/10.17504/protocols.io.j87crzn



James E Thornton Jr¹

¹Hurwitz Lab

Metafunc Course 2017



James E Thornton Jr

Hurwitz Lab

OPEN  ACCESS



DOI: dx.doi.org/10.17504/protocols.io.j87crzn

Protocol Citation: James E Thornton Jr 2017. Assembly with Megahit. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.j87crzn>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

Created: October 10, 2017

Last Modified: March 28, 2018

Protocol Integer ID: 8191

Abstract

Co-assembly using Megahit.



1 Log into the HPC.

Command

```
$ ssh hpc  
$ ocelote
```

2 From your home directory, open .bashrc file for editing.

Command

```
$ nano .bashrc
```

Note

Remember, you are already in your home directory after logging into ocelote.

3 Input the following line into your .bashrc file:

Command

This will allow us to execute tools found in /rsgroups/bh_class/bin without specifying the path name.

```
export PATH=/rsgroups/bh_class/bin:$PATH
```

4 Save and close the .bashrc file



- 5 Move into your project directory.

Command

```
$ cd /rsgrps/bh_class/username
```

- 6 Create a directory for assembly output. Then move into that directory.

Command

```
$ mkdir assembly  
$ cd !
```

- 7 Make directories for standard out and standard error.

Command

```
mkdir std-out std-err
```

- 8 Before we continue, determine if you have single end or paired end files. If you have two files per SRR number, you have paired end reads. Otherwise, you have single end reads.

1. If you have single end reads proceed to step 5.
2. If you have paired end reads, skip to step 6.

- 9 Assembly script for SINGLE END FILES

Create a script called run-assembly.sh



Command

```
#!/bin/bash
#PBS -W group_list=bh_class
#PBS -q windfall
#PBS -l select=1:ncpus=20:mem=40gb
#PBS -l pvmem=38gb
#PBS -l walltime=24:00:00
#PBS -l cput=48:00:00
#PBS -M netid@email.arizona.edu
#PBS -m bea
FASTQ_DIR='/rsgtps/bh_class/username/fastq'
ASSEM_DIR='/rsgtps/bh_class/username/assembly'
MIN_CONTIG_LEN=500
OUT_DIR='/rsgtps/bh_class/username/assembly/megahit-out'
cd $ASSEM_DIR
SINGLES=`ls $FASTQ_DIR/*.fastq | python -c 'import sys; print`
```

Note

OUT_DIR does NOT need to be created prior to running this script. Megahit will make the directory on its own.

10 Assembly script for PAIRED END FILES

Create a script called run-assembly.sh

**Command**

```
#!/bin/bash
#PBS -W group_list=bh_class
#PBS -q windfall
#PBS -l select=1:ncpus=20:mem=40gb
#PBS -l pvmem=38gb
#PBS -l walltime=24:00:00
#PBS -l cput=48:00:00
#PBS -M netid@email.arizona.edu
#PBS -m bea
FASTQ_DIR='/rsgrps/bh_class/username/fastq'
ASSEM_DIR='/rsgrps/bh_class/username/assembly'
MIN_CONTIG_LEN=500
OUT_DIR='/rsgrps/bh_class/username/assembly/megahit-out'
cd $ASSEM_DIR
R1s=`ls $FASTQ_DIR/*_1.fastq | python -c 'import sys; print
```

Note

OUT_DIR does NOT need to be created prior to running this script. Megahit will make the directory on its own.

- 11 Run the assembly:

Command

```
$ chmod +x run-assembly.sh
$ qsub -e std-err/ -o std-out/ run-assembly.sh
```

- 12 You can check the status of your job with the following command:

**Command**

```
$ qstat -u username
```

Note

Job runtime will vary depending on the size of your dataset.

- 13 Upon job completion, get assembly statistics using MetaQuast on CyVerse.

Protocol

NAME

Assembly Stats with MetaQuast

CREATED BY

James E Thornton Jr

PREVIEW

- 13.1 Go to <https://user.cyverse.org/>
<https://user.cyverse.org/>
- 13.2 Click "Sign Up" to create an account.
- 13.3 After account creation go back to <https://user.cyverse.org/> and login with your account.
<https://user.cyverse.org/>
- 13.4 Launch the discovery environment.



Discovery Environment

Use hundreds of bioinformatics apps and manage data in the interface

13.5 Click the "Data" button found on the left. Navigate to your user folder.

The screenshot shows the CyVerse Discovery Environment interface. On the left, there are three main buttons: 'Data' (cloud icon), 'Apps' (grid icon), and 'Analyses' (monitor icon). The 'Data' button is selected. The main panel displays a file browser for user 'jetjr'. The navigation pane on the left shows folders: 'jetjr', 'Community Data', 'Shared With Me', 'Trash', and 'Favorites'. The main pane shows a list of files and folders with columns for 'Name' and 'Last Modified'.

Name	Last Modified
App_data	2017 May 26 12:21
HOT	2015 Sep 22 07:36
HOT2	2015 Sep 22 07:51
Project	2015 Oct 30 05:39
analyses	2015 Sep 22 07:51
archive	2017 May 3 12:29
assembly_tutorial	2015 Oct 12 20:15
coge_data	2016 May 24 12:41
infantgut_3do	2016 Feb 20 10:21
neutropenicfever	2016 Apr 26 10:32
obese_lean	2015 Oct 26 11:20
outreach	2017 Feb 22 07:55

Displaying 1 - 13 of 17

5

13.6 Click "Upload" > "Simple Upload From Desktop"

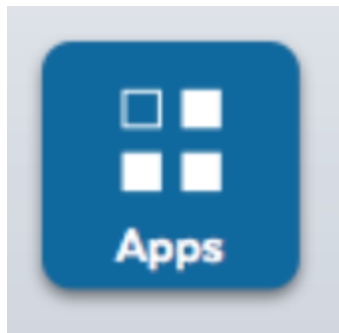
13.7 Upload your final.contigs.fa file generated from Megahit.

Note

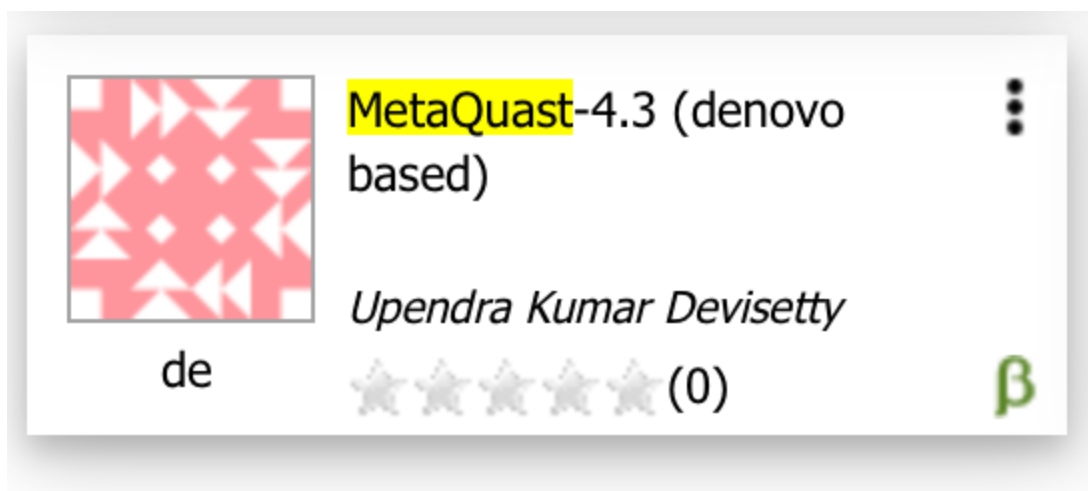
Important: You must scp your contigs to your local machine before you can upload.

\$ scp username@sftp.hpc.arizona.edu:/rsgrps/bh_class/username/assembly/megahit-out/final.contigs.fa .

13.8 Once your upload is complete, click on the "Apps" button found on the left.



13.9 Search for "MetaQuast". Click on MetaQuast-4.3 (denovo based)





- 13.10 Under the "Fasta file(s)" tab, select the newly uploaded final.contigs.fa file. This is the only parameter that needs to change. Click "Launch Analysis".
- 13.11 Once MetaQuast is complete (email notification), navigate to the output found in the "analyses" folder in your data storage.
- 13.12 Download the "report.html" file found in the MegaQuast output folder.
- 13.13 Open the report.html file to see a summary of assembly statistics.