

Assembling Viral Metagenomic Data with SPAdes (Cyverse)

Benjamin Bolduc

Abstract

Assembling *actual* reads from the [Ocean Sampling Day \(2014\)](#) using SPAdes, an assembler implemented in [Cyverse](#).

Citation: Benjamin Bolduc Assembling Viral Metagenomic Data with SPAdes (Cyverse). **protocols.io**
dx.doi.org/10.17504/protocols.io.evzbe76

Published: 22 Apr 2016

Guidelines

This is part of a larger protocol *Collection* that involves the end-to-end processing of raw viral metagenomic reads obtained from a sequencing facility to assembly and analysis using Apps (i.e. tools) developed by iVirus and implemented within the Cyverse cyberinfrastructure.

Assembly is just one part of the big picture!

Before start

To run this protocol, users must first [register](#) for Cyverse account. All data (both inputs and outputs) are available within Cyverse's data store at `/iplant/home/shared/iVirus/ExampleData/`

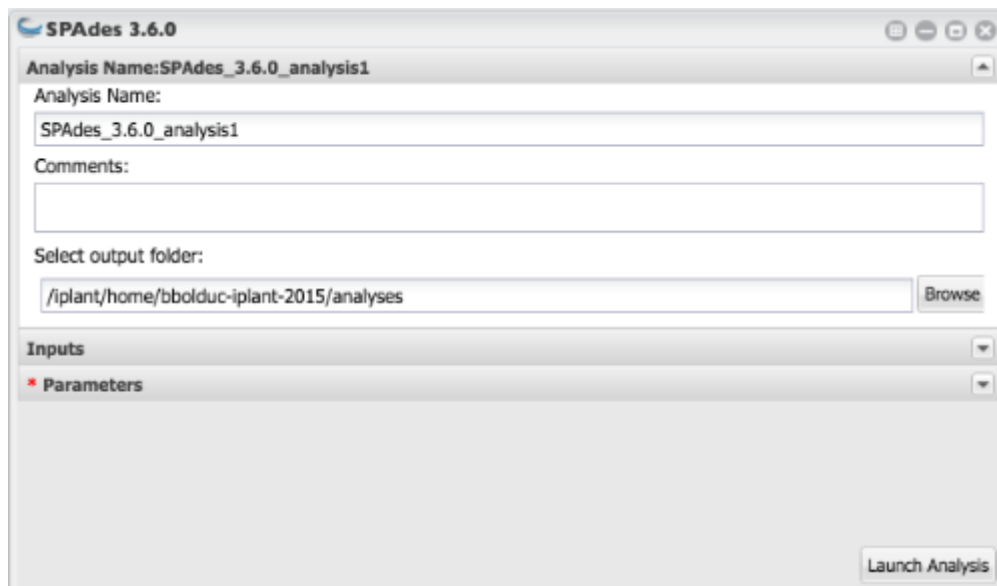
Protocol

SPAdes

Step 1.

Open SPAdes

Open SPAdes from "Apps."



SPAdes

Step 2.

Select Inputs

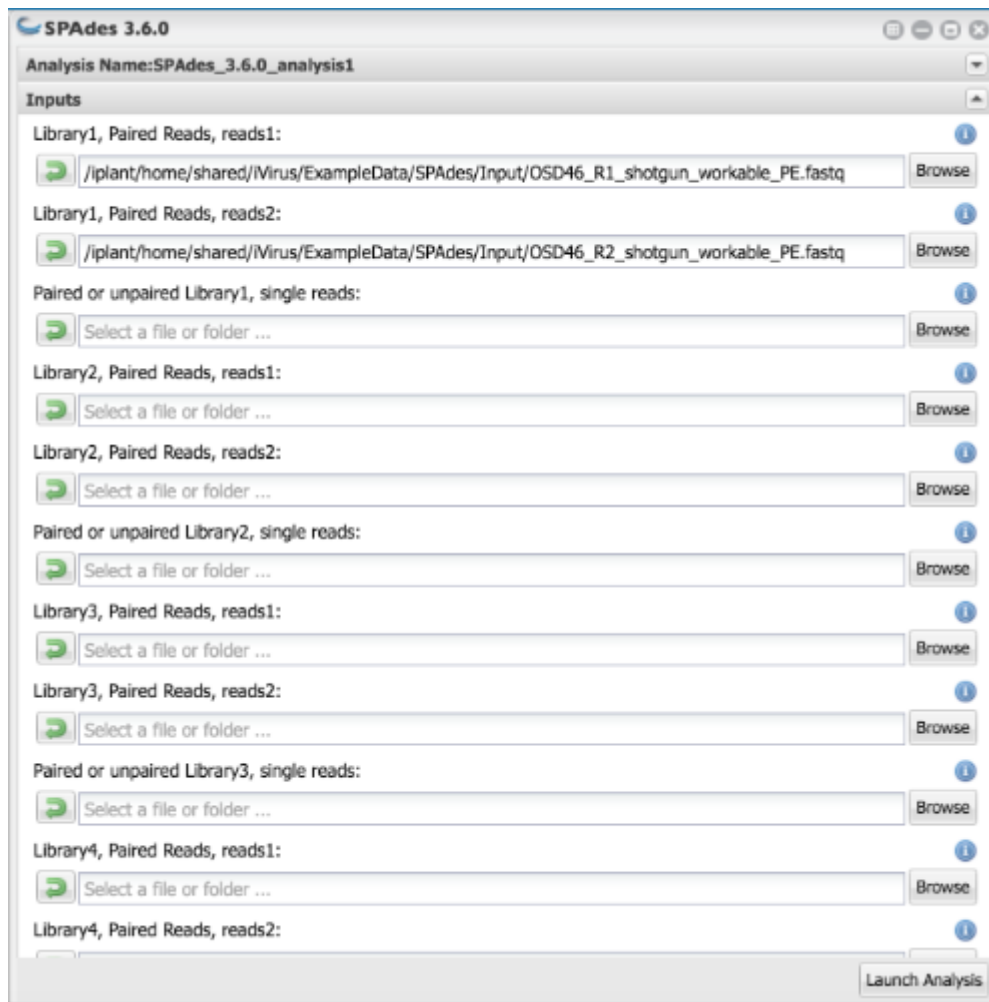
Select the 'Inputs' tab.

For **Library1, Paired Reads, reads1**

- Navigate to *Community Data --> iVirus --> ExampleData --> SPAdes --> Input*. Select *OSD46_R1_shotgun_workable_PE.fastq* Alternatively, copy-and-paste the location: */iplant/home/shared/iVirus/ExampleData/SPAdes/Inputs* into the navigation bar and select the fastq file.

For **Library2, Paired Reads, reads2**

- Navigate to *Community Data --> iVirus --> ExampleData --> SPAdes --> Input*. Select *OSD46_R2_shotgun_workable_PE.fastq* Alternatively, copy-and-paste the location: */iplant/home/shared/iVirus/ExampleData/SPAdes/Inputs* into the navigation bar and select the fastq file.



📌 NOTES

Benjamin Bolduc 20 Apr 2016

The read files used in this example can be taken directly from the iVirus' SPAdes input directory on Cyverse, or they can be generated using *raw* reads from the OSD website and processed using Trimmomatic.

SPAdes

Step 3.

Select Parameters

The default parameters will be sufficient for this example. Of note is the change in 'Kmer setting(s)' where we have changed the default (35) to a mixture of 21,33,55. We have found this produces good viral assemblies, though your mileage may vary.

Also to note are the fact that the 'Input' files are directly dependent on the 'Parameters.' In this example we selected 2 paired end files. *Each of these* are 'paired reads' because they have partners, and the 'pairing format, library 1' is 'paired end.' With other data, users can mix-and-match paired and unpaired reads for up to 5 libraries.



SPAdes

Step 4.

Launch Analysis

Run the job! Depending on the dataset, this could take hours or days. This sample dataset should only take a few hours.

NOTES

Benjamin Bolduc 20 Apr 2016

There is a "high-mem" version of SPAdes that may be useful if SPAdes fails *and the reason for failure is insufficient memory*. **Not all SPAdes failures are due to memory.**

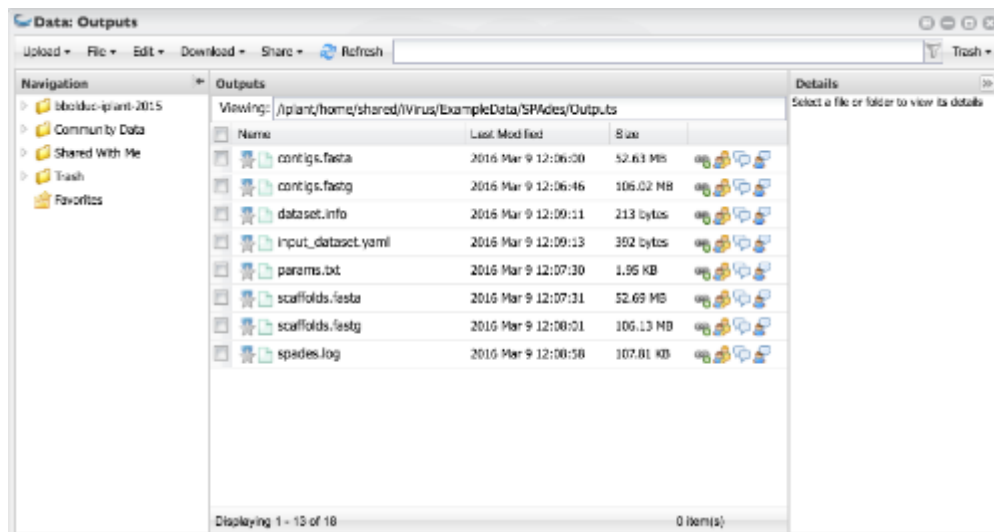
SPAdes

Step 5.

Results

Expect results can be found in the SPAdes' 'Output' directory.

An abbreviated output of important files generated during assembly. The most important files are *contigs.fasta* and *scaffolds.fasta*. They contained the assembled sequences (i.e. contigs). You should save this file



✓ EXPECTED RESULTS

Full output when 3 different Kmers are selected...

