Identifying Viral Sequences Using VirSorter (Cyverse)

Benjamin Bolduc

Abstract

Identifying *putative* viral sequences from SPAdes-assembled data from the Ocean Sampling Day (2014) metagenomic datasets using VirSorter.

Citation: Benjamin Bolduc Identifying Viral Sequences Using VirSorter (Cyverse). **protocols.io** dx.doi.org/10.17504/protocols.io.ev2be8e
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.

Finding putative viral sequences is one spring board for viral metagenomic analysis.

Before start

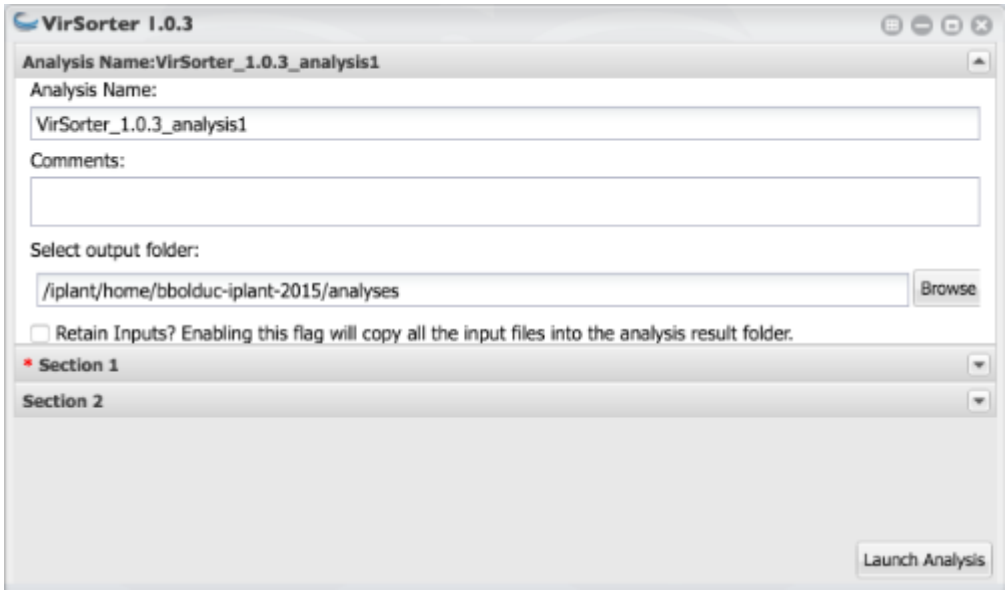
To run this protocol, users must first register for Cyverse account. All data (both inputs and outputs) are available within Cyverse's data store at /iplant/home/shared/iVirus/ExampleData/

Protocol

VirSorter
Step 1.

Open VirSorter

Open VirSorter from 'Apps'



ANNOTATIONS

German Bonilla 04 Jul 2017

It seems that the new layout of protocols.io messed up all iVir instructions...

Is there any other place where these instructions can be accessed?

VirSorter

Step 2.

Select Inputs

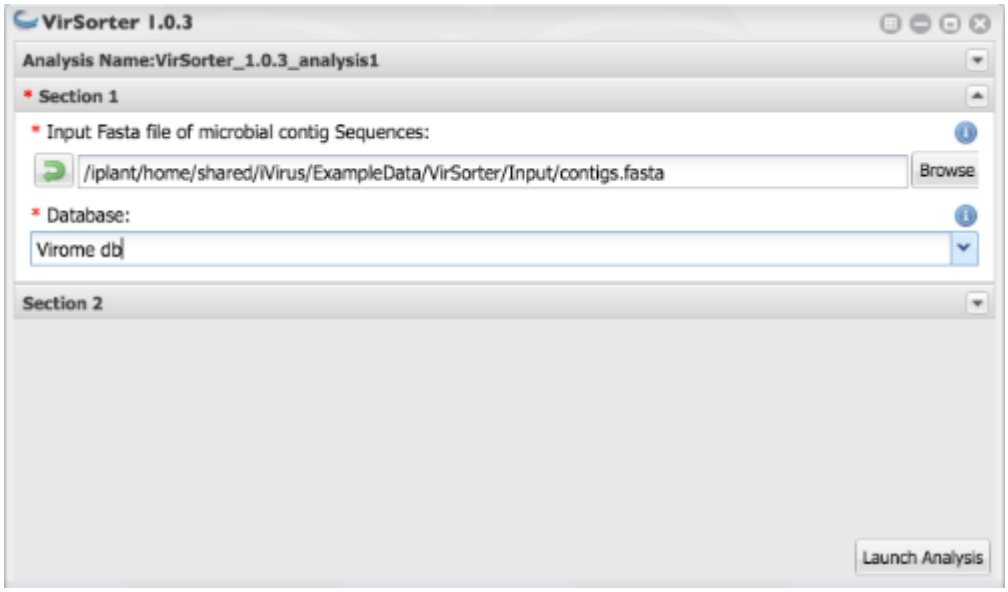
Select the 'Section 1' tab.

For **Input Fasta file of microbial contig Sequences**

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

For **Database**

- Select *Virome*. There are only two databases to select. Virome and RefSeq.



VirSorter

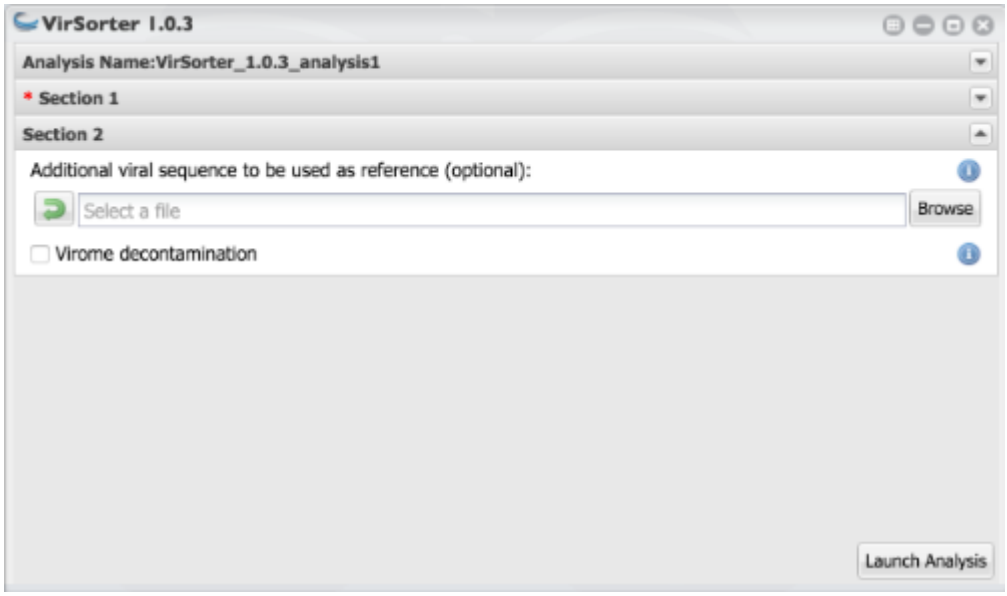
Step 3.

Select Parameters

Select the 'Section 2' tab.

The default options in this section will be sufficient.

Additional viral sequence to be used as reference: Here users can select a fasta file that includes viruses that may not be in either 'Virome' or 'RefSeq' databases. For example, unpublished viral sequences used internally in a lab.
Virome Decontamination: Virsorter was designed to screen for viral sequences within *microbial* datasets. Part of its methodology identifies the 'non-viral' component and uses that as background to improve differentiation between viral and non-viral sequences. If most and/or all of the sequences are potentially viral, select this option. VirSorter will use a 'generic' cellular dataset as background.



VirSorter

Step 4.

Launch Analysis

Run the job!

VirSorter can take minutes, hours or days depending on the size of your dataset. For this example it should take an afternoon.

VirSorter

Step 5.

Results

Expected results can be found from the 'Outputs' directory of VirSorter.

Navigate to *Community Data* --> *iVirus* --> *ExampleData* --> VirSorter --> *Output*. Alternatively, copy-and-paste the location: /iplant/home/shared/iVirus/ExampleData/VirSorter/Output into the navigation bar.