Applying vContact to Viral Sequences and Visualizing the Output (Cyverse)

Benjamin Bolduc

Abstract

A collection of protocols designed to guide the user in processing a viral metagenome from raw sequence data to assembly, and subsequent analysis. The user uses *actual* reads from [Ocean Sampling Day (2014)](#) and processes them entirely within Cyverse, a NSF-supported cyberinfrastructure.

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.

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/

1. Download and install [Java JDK 8](#)
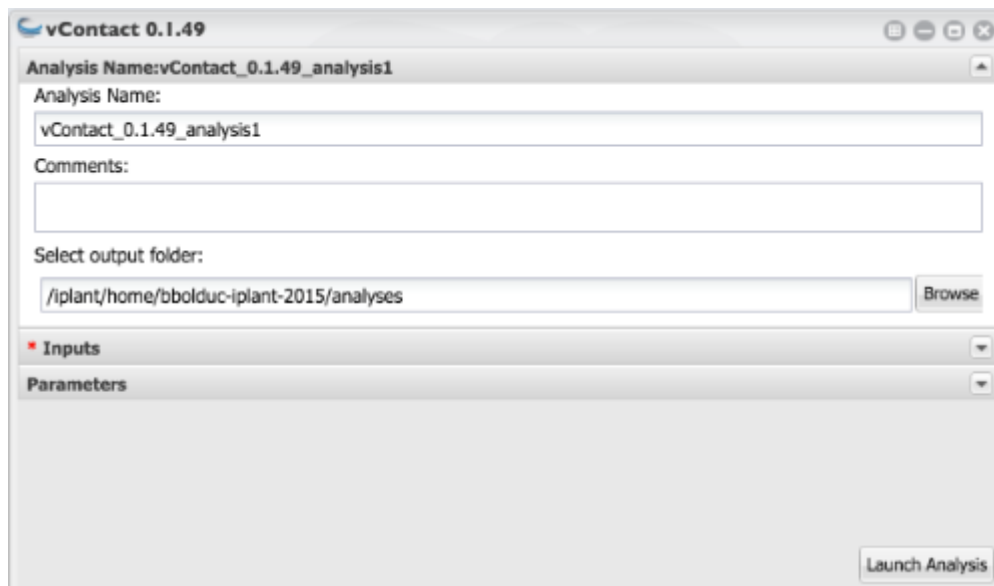2. Download and install [Cytoscape 3.x](#)

Protocol

Affiliating contigs through their shared proteins
Step 1.

Open vContact

Open vContact from 'Apps'

Affiliating contigs through their shared proteins

Step 2.

Select Inputs

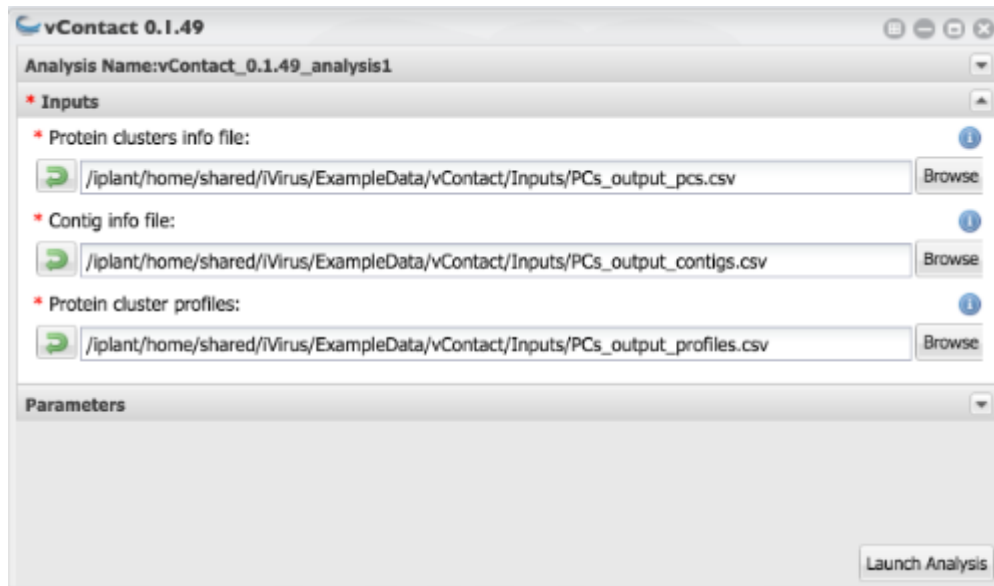Select the 'Inputs tab.

For **Protein clusters info file**:

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

For the **Contig info file**:

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

For **Protein cluster profiles**:

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

Affiliating contigs through their shared proteins

Step 3.

Select Parameters

Select the 'Parameters' tab.

The default options will suffice for this example. Consult the relevant documentation for what each of these options mean.

Affiliating contigs through their shared proteins

Step 4.

Launch Analysis

Run the job!

vContact can take minutes to hours to the better part of a day to complete.

Step 5.

Results

vContact will generate a results folder with network files (*.ntw), contig clustering information (cc_*) and modules (mod_*). The network files can be imported into Cytoscape to visualize the modules and the contig clusters.

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

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

Step 6.

Open Cytoscape

Open Cytoscape *on your local machine*.

Step 7.

Locate and Select Network File

- If a 'splash window' appears, select 'Start New Session - From Network File...'
- If the window doesn't appear, go to File -> Import -> Network -> File...

Select the contig *.ntw (typically *cc_sig1.0_mcl2.o.ntw).

Step 8.

Import Network File

1. Select "Advanced Options" and select the appropriate Delimiter, in this case "SPACE." and click "OK."
   - At this point you can change the "Default Interaction" to something more meaningful, or keep as is.
   - This changes the single column import into 3 (there might be one hiding on the right)
2. Click on "Column 1" and under *Meaning*, select *Source Node* (little green button).
3. Click on "Column 2" and under *Meaning*, select *Target Node* (red bullseye).
4. Click on "Column 3" and under *Meaning*, select *Edge Attribute* (purple file).
5. Select "Ok." One this happens, it might take a while to load the network.

Step 9.

Results

Depending on the size of your network, Cytoscape might not automatically create a *View* for the network. Our example case is small enough so it should automatically create one. However, real data often has 100s, 1000s, 10s of 1000s of nodes and can be memory intensive.

If your data is large, you can still visualize the network. A popup will appear, "Create Network Views?" Select "Ok." Once finished, the network view will be *roughly* ordered by cluster size!