Quality Control of Reads Using Trimmomatic (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.

Quality trimming of reads is important!

Citation: Benjamin Bolduc Quality Control of Reads Using Trimmomatic (Cyverse). **protocols.io** dx.doi.org/10.17504/protocols.io.gvybw7w
Published: 31 Dec 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.

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/

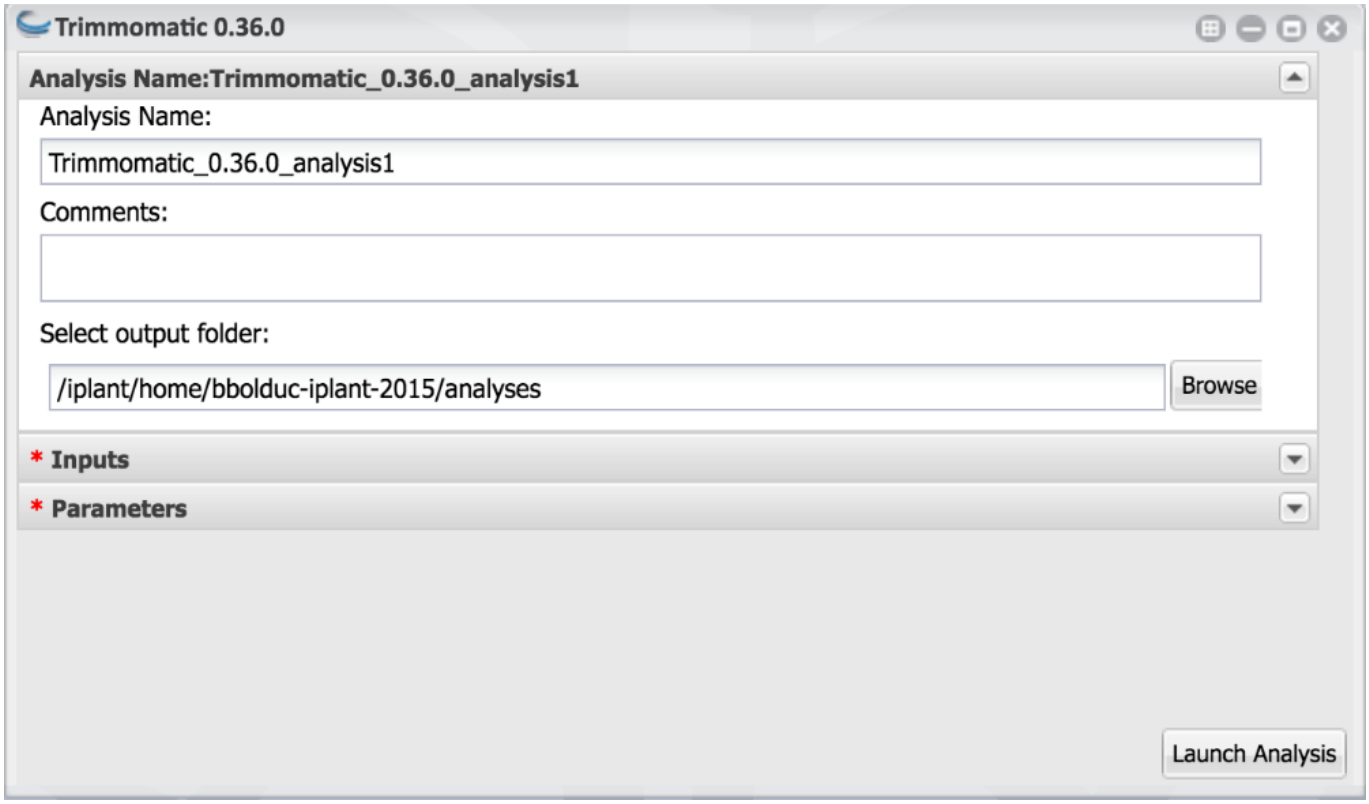
Protocol

Quality control of reads

Step 1.

Open Trimmomatic

Open Trimmomatic 0.36.0 from 'Apps'



Quality control of reads

Step 2.

Select Inputs

Select the 'Inputs tab.

For **Reads file (for paired or unpaired)**:

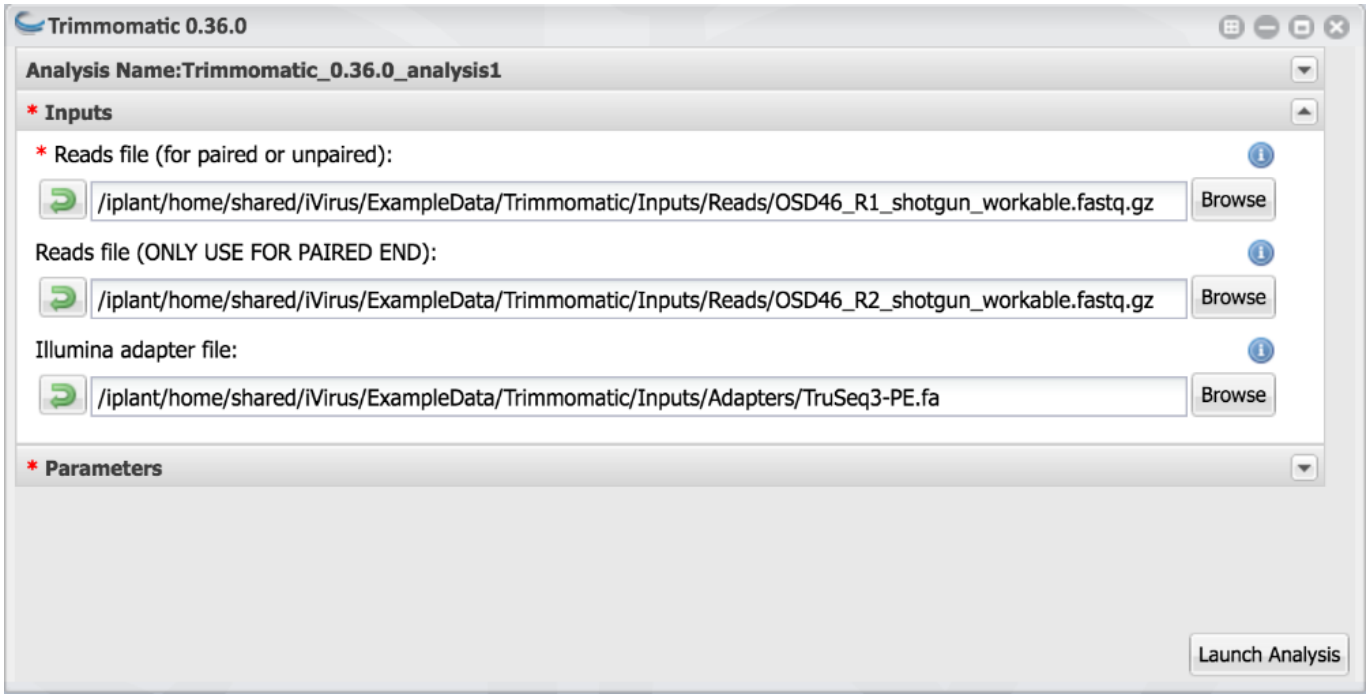
- Navigate to *Community Data --> iVirus --> ExampleData --> Trimmomatic --> Inputs --> Reads*. Select *OSD46_R1_shotgun_workable.fastq.gz* Alternatively, copy-and-paste the location: /iplant/home/shared/iVirus/ExampleData/Trimmomatic/Inputs/Reads into the navigation bar and select the fastq.gz file.

For **Reads file (ONLY USE FOR PAIRED END)**: Only use this file if it's read pair 2. Don't try to double-up two single-end reads.

- Navigate to *Community Data --> iVirus --> ExampleData --> Trimmomatic --> Inputs --> Reads*. Select*OSD46_R2_shotgun_workable.fastq.gz* Alternatively, copy-and-paste the location: /iplant/home/shared/iVirus/ExampleData/Trimmomatic/Inputs/Reads into the navigation bar and select the fastq.gz file.

For **Illumina adapter file**:

- Navigate to *Community Data --> iVirus --> ExampleData --> Trimmomatic --> Inputs --> Adapters*. Select *TruSeq3-PE.fa* Alternatively, copy-and-paste the location: /iplant/home/shared/iVirus/ExampleData/Trimmomatic/Inputs/Adapters into the navigation bar and select the fasta file.

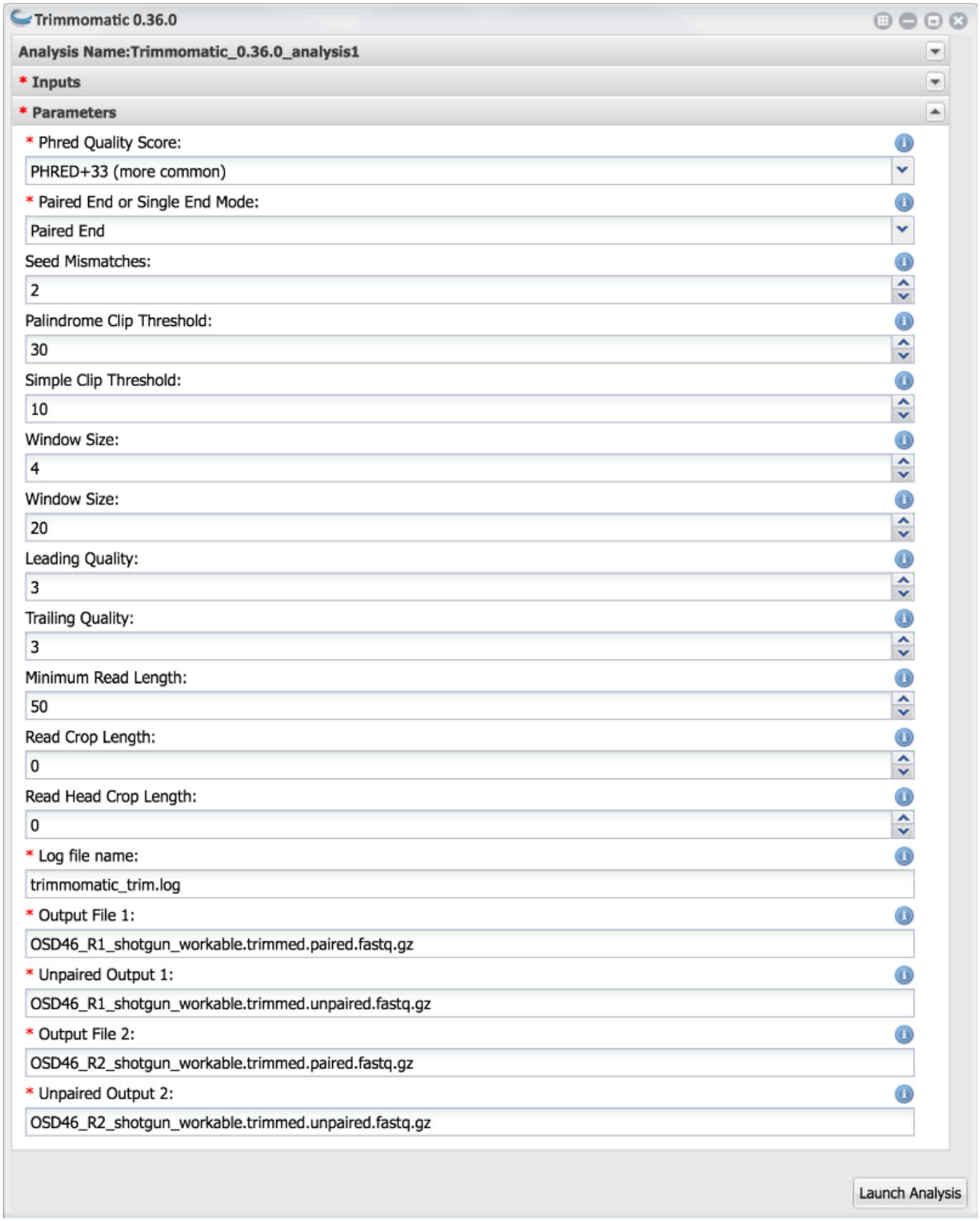


Step 3.

Select Parameters

Select the 'Parameters' tab.

The default options will suffice for this example.



NOTES

Benjamin Bolduc 30 Dec 2016

TIP: A full explanation of all the parameters can be found at the Trimmomatic web site. These are the defaults that work well for most data.

Benjamin Bolduc 30 Dec 2016

PRO TIP: To conserve disk space, Trimmomatic will recognize the gzip extension (*.gz), so if you want the output to be compressed, add ".gz" to the extension and it will automatically compress the results!

Quality control of reads

Step 4.

Launch Analysis

Run the job!

Trimmomatic can take minutes to hours to complete.

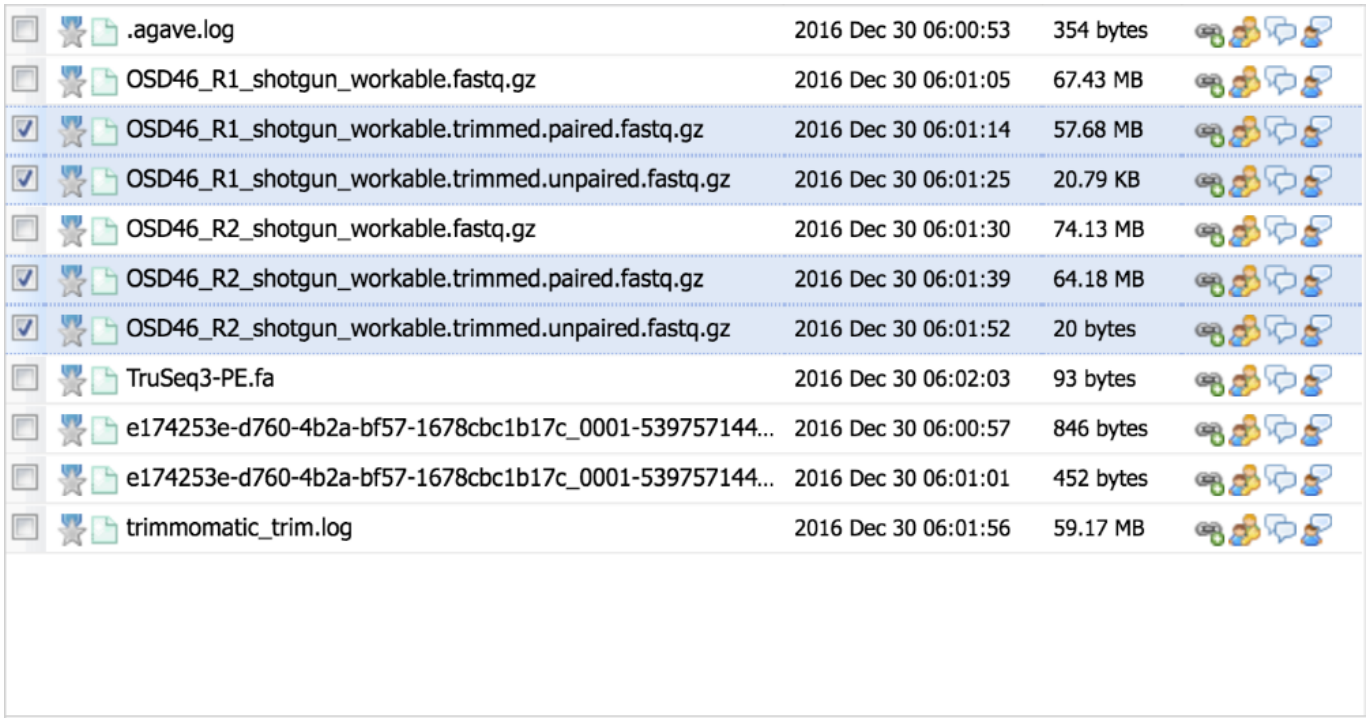
Quality control of reads

Step 5.

Results

Trimmomatic will generate numerous output files. 2 paired files where both forward and reverse survived read trimming, 2 unpaired files where their pair/mate did not survive, a log file, the inputs and an agave.log file. The most important are the forward and reverse reads surviving, and the unpaired versions.

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



NOTES

Benjamin Bolduc 30 Dec 2016

TIP: In general, most of the read data should be in the 'paired' forward and reverse reads. *If its not*, it may indicate poorer quality input data and/or issues with library prep/DNA quality (the 'material' that went into the library prep). If that's true, it's not the end of the world. Good assemblies can still be generated using the *now* high-quality reads that remain.