

Dec 30, 2016

Quality Control of Reads Using Trimmomatic (Cyverse)

 Forked from [Quality Control of Reads Using Trimmomatic \(Cyverse\)](#).

DOI

dx.doi.org/10.17504/protocols.io.gvybw7w

Benjamin Bolduc¹

¹The Ohio State University

VERVE Net

Sullivan Lab

1 more workspace



Benjamin Bolduc

The Ohio State University

OPEN  ACCESS



DOI: dx.doi.org/10.17504/protocols.io.gvybw7w

Protocol Citation: Benjamin Bolduc 2016. Quality Control of Reads Using Trimmomatic (Cyverse). **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.gvybw7w>

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: December 30, 2016

Last Modified: March 26, 2018

Protocol Integer ID: 4760

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!

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.

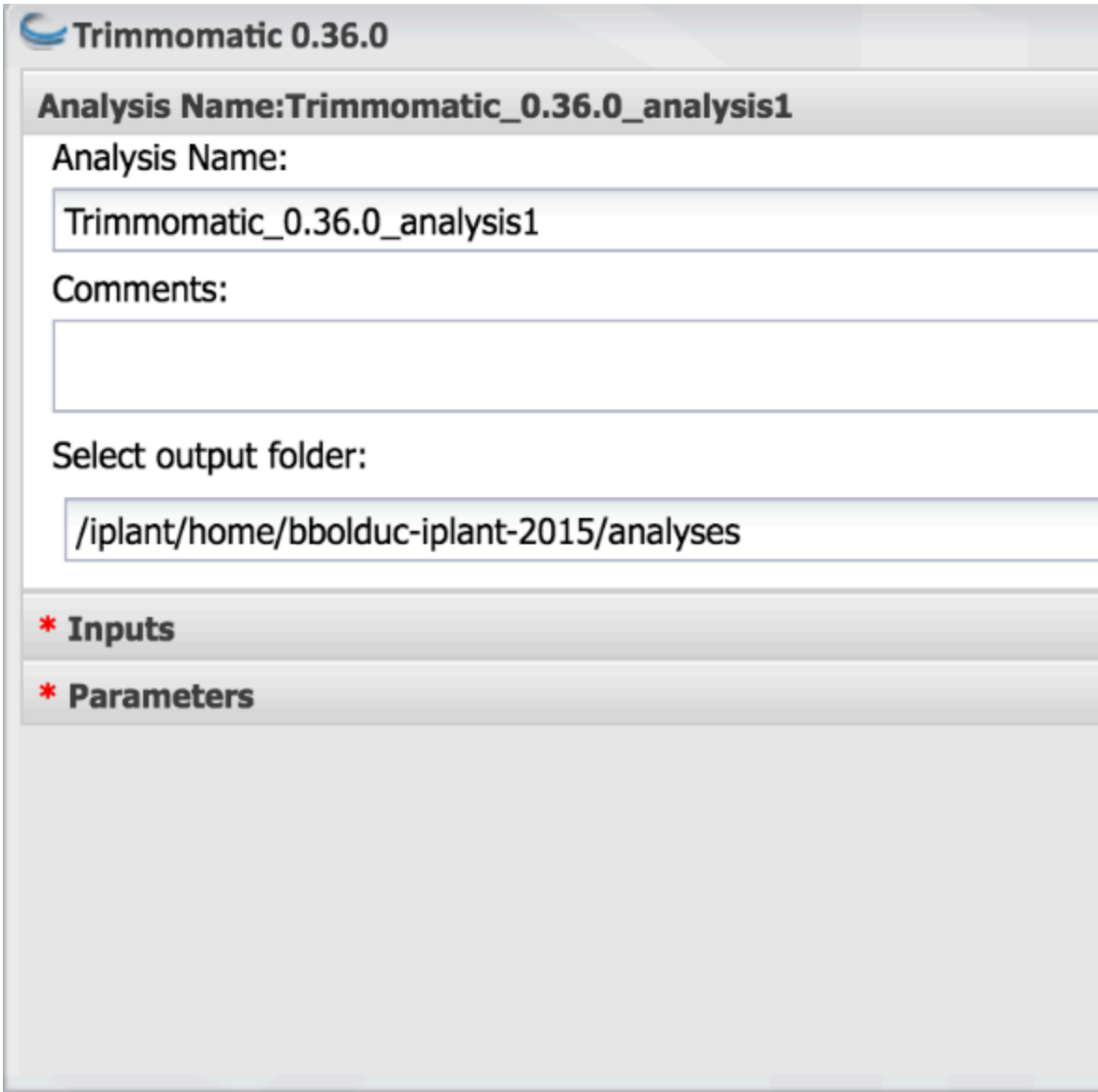
Before start

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/`

Quality control of reads

1 **Open Trimmomatic**

Open Trimmomatic 0.36.0 from 'Apps'



Trimmomatic 0.36.0

Analysis Name: Trimmomatic_0.36.0_analysis1

Analysis Name:
Trimmomatic_0.36.0_analysis1

Comments:

Select output folder:
/iplant/home/bbolduc-iplant-2015/analyses

* **Inputs**

* **Parameters**

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.

The screenshot shows the Trimmomatic 0.36.0 web interface. The analysis name is 'Trimmomatic_0.36.0_analysis1'. Under the 'Inputs' section, there are three input fields, each with a green arrow icon on the left:

- * Reads file (for paired or unpaired):
- Reads file (ONLY USE FOR PAIRED END):
- Illumina adapter file:

Below the inputs is a section for 'Parameters' which is currently empty.

3 **Select Parameters**

Select the 'Parameters' tab.

The default options will suffice for this example.

Trimmomatic 0.36.0

Analysis Name: Trimmomatic_0.36.0_analysis1

*** Inputs**

*** Parameters**

*** Phred Quality Score:**

*** Paired End or Single End Mode:**

Seed Mismatches:

Palindrome Clip Threshold:

Simple Clip Threshold:

Window Size:

Window Size:

Leading Quality:

Trailing Quality:

Minimum Read Length:

Read Crop Length:

Read Head Crop Length:

*** Log file name:**

*** Output File 1:**

*** Unpaired Output 1:**

*** Output File 2:**

*** Unpaired Output 2:**



```
OSD46_R2_shotgun_workable.trimmed.unpaired.fastq.gz
```

Note

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.

Note

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

4 **Launch Analysis**























Run the job!

Trimmomatic can take minutes to hours to complete.

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.

<input type="checkbox"/>	  .agave.log	2
<input type="checkbox"/>	  OSD46_R1_shotgun_workable.fastq.gz	2
<input checked="" type="checkbox"/>	  OSD46_R1_shotgun_workable.trimmed.paired.fastq.gz	2
<input checked="" type="checkbox"/>	  OSD46_R1_shotgun_workable.trimmed.unpaired.fastq.gz	2
<input type="checkbox"/>	  OSD46_R2_shotgun_workable.fastq.gz	2
<input checked="" type="checkbox"/>	  OSD46_R2_shotgun_workable.trimmed.paired.fastq.gz	2
<input checked="" type="checkbox"/>	  OSD46_R2_shotgun_workable.trimmed.unpaired.fastq.gz	2
<input type="checkbox"/>	  TruSeq3-PE.fa	2
<input type="checkbox"/>	  e174253e-d760-4b2a-bf57-1678cbc1b17c_0001-539757144...	2
<input type="checkbox"/>	  e174253e-d760-4b2a-bf57-1678cbc1b17c_0001-539757144...	2
<input type="checkbox"/>	  trimmomatic_trim.log	2

Note

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.