

ShigaTyper

ShigaTyper is a quick and easy tool designed to determine *Shigella* serotype using Illumina paired end reads with low computation requirement. It has been tested on Ubuntu 16.04.3 LTS and Ubuntu 18.04 LTS guest additions on VMware Player version 14.1.1 in a Windows 7 and Virtual Box version 5.2.4 in a Windows 10 operating system. If you are not familiar with Linux system or WGS analysis, please refer to the ShigaTyper user tutorial, where instructions are included for system setup and how to run the analysis.

Dependencies

All can be installed via Anaconda

1. Python 3.6 or later
2. Jupyter Notebook (install from Anaconda with Python 3.6 or later)
3. fastp version 0.12.2 or later
4. minimap2 version 2.9 or later
5. samtools version 1.7 or later
6. bcftools version 1.7 or later
7. papermill version 0.12.5 or later

Accepted file formats

1. Illumina paired-end reads in fastq.gz or fastq format
2. Assembled genome in fasta format (note: not validated)

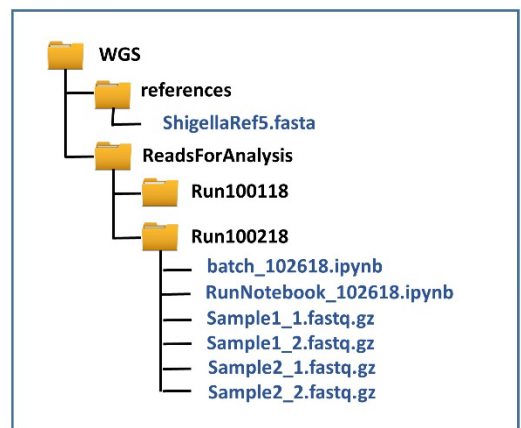
Run ShigaTyper

Download the following 3 files and place them following the file organization shown in the figure below:

1. batch_010819.ipynb
2. RunNotebooks_010819.ipynb
3. ShigellaRef6.fasta

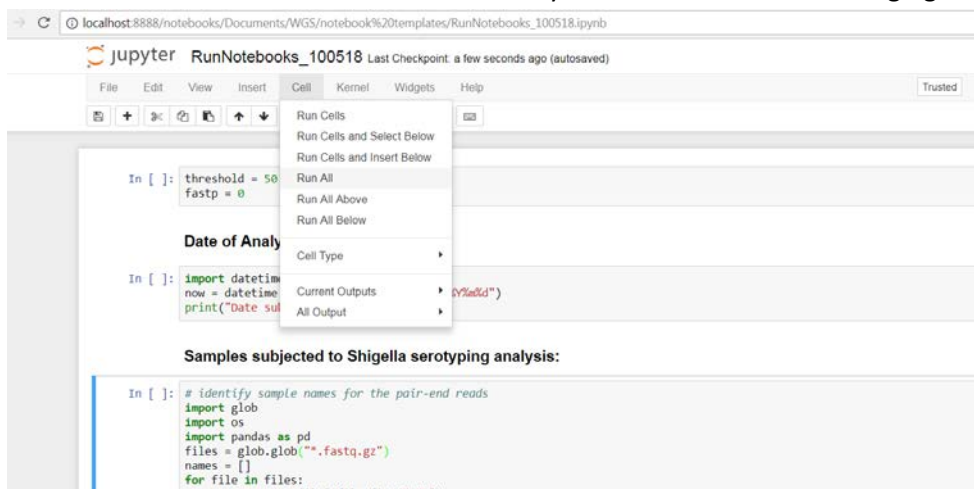
Note: if you change the reference file path or name you will have to edit the directory to reference (ShigellaRef) in “RunNotebooks_010819.ipynb”. Currently it is set as

ShigellaRef = “.././references/ShigellaRef6.fasta”
(in the 3rd cell).

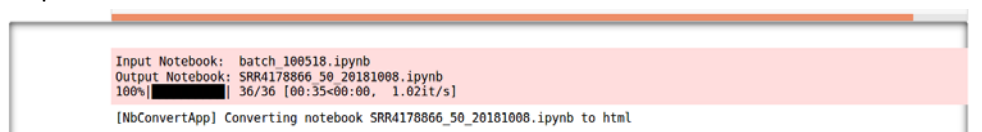


[optional] if you want to omit QC inspection, in the first cell of RunNotebooks_010819.ipynb, change from “fastp = 0” to “fastp = 1”.

1. Open Terminal and switch directory to the folder containing WGS reads, or its parent folder.
2. Type “jupyter notebook” to turn on Jupyter Notebook in the default browser (Firefox in Ubuntu).
3. Navigate to the folder containing the jupyter notebooks and fastq.gz files to be analyzed.
4. Click on “RunNotebooks_010819.ipynb” to open it in a new tab.
5. Click on “Cell” then “Run All” to start the analysis as shown in the following figure:



- 5.1. The notebook will automatically run to finish. Samples analyzed before in this notebook will be erased and replaced by results from the new samples. You can see the progress for each of the sample:



- 5.2. A html report is automatically generated for each of the sample.
- 5.3. When the analysis of all samples is complete, a summary table is generated at the end of RunNotebook_010819.ipynb as below:

```
Date of analysis: 20181008
Threshold level for gene coverage: 50 %
Quality inspection conducted: Yes
1 samples were analyzed:
```

Sample	Size (MB)	Prediction
SRR4178866	170.9	Shigella flexneri serotype 6

6. Click “File” then “Save” to save the results. You can change the filename of the notebook (for example, “ShigaTyper_ShigellaSamples_012519”) by clicking on the title, “RunNotebooks_010819”, on the top next to “jupyter”.
7. To download a html version of the summary, click on “File”, “Download as”, then “html”.
8. Close Jupyter Notebook.
 - 8.1. Click on “File” then “Close and Halt” to close the notebook.
 - 8.2. Click “Logout” on the upper right corner in the main page. Quit browser.
 - 8.3. Enter “Ctrl + c” in Terminal then type “y” upon prompt to quit Jupyter Notebook.