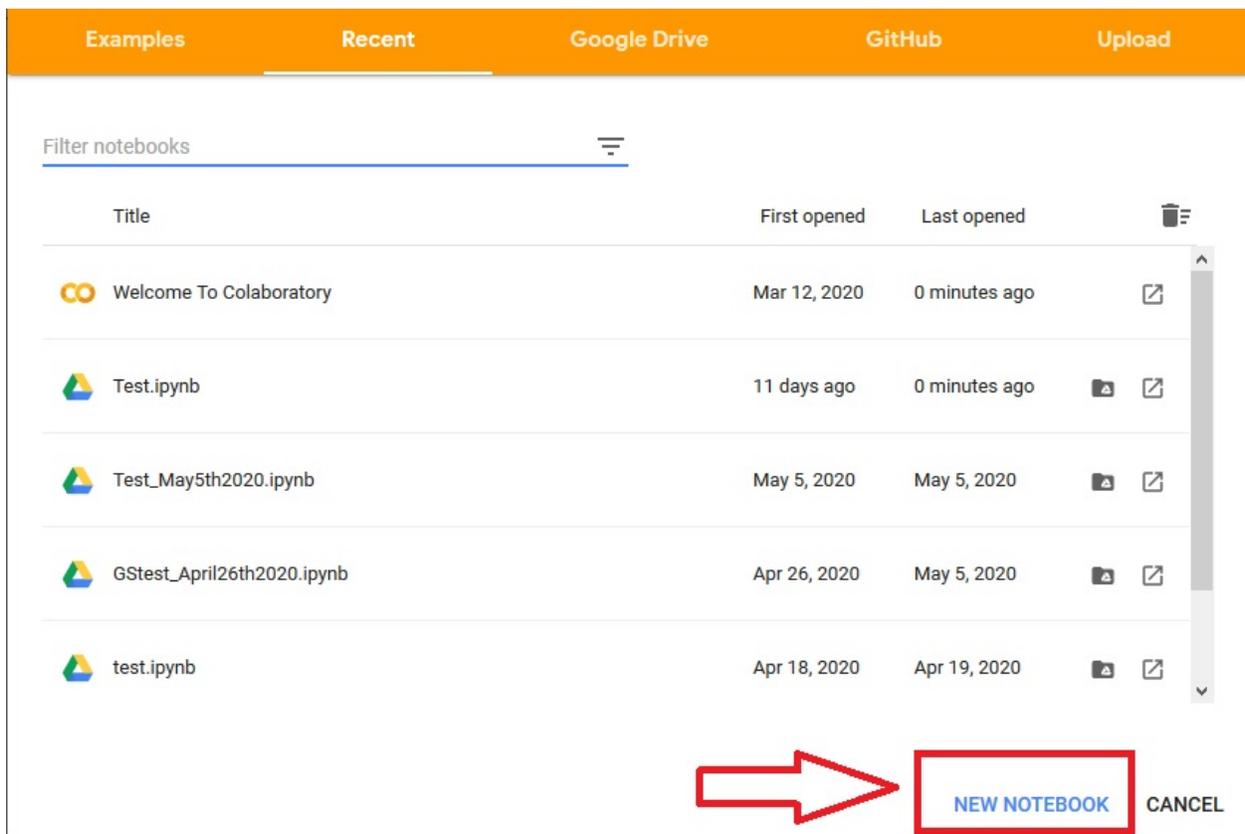


## Instructions to run the Pipeline:

**Github page:** <https://github.com/genomesolver/CSPpipeline>

1. Open Google Collab: <https://colab.research.google.com/notebooks/welcome.ipynb> and Click on New Notebook. Sign into your google account. The output from the python script will be stored in your google drive.

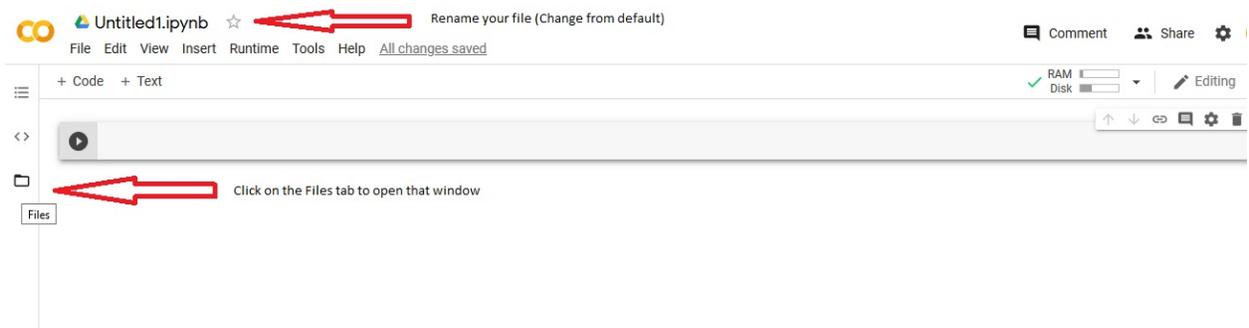


The screenshot shows the Google Colab interface with a list of notebooks. The interface has a top navigation bar with tabs for 'Examples', 'Recent', 'Google Drive', 'GitHub', and 'Upload'. Below the navigation bar is a search bar labeled 'Filter notebooks'. The main content area displays a table of notebooks with columns for 'Title', 'First opened', 'Last opened', and a set of icons (trash, share, refresh). The notebooks listed are:

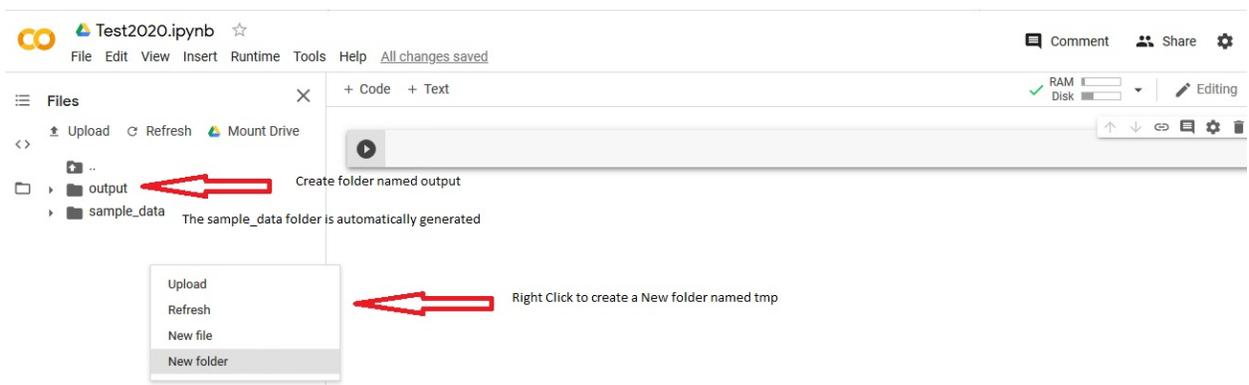
Title	First opened	Last opened	Icons
Welcome To Colaboratory	Mar 12, 2020	0 minutes ago	Share icon
Test.ipynb	11 days ago	0 minutes ago	Share icon, Refresh icon
Test_May5th2020.ipynb	May 5, 2020	May 5, 2020	Share icon, Refresh icon
GStest_April26th2020.ipynb	Apr 26, 2020	May 5, 2020	Share icon, Refresh icon
test.ipynb	Apr 18, 2020	Apr 19, 2020	Share icon, Refresh icon

At the bottom right of the interface, there is a 'NEW NOTEBOOK' button highlighted with a red box and a red arrow pointing to it, and a 'CANCEL' button next to it.

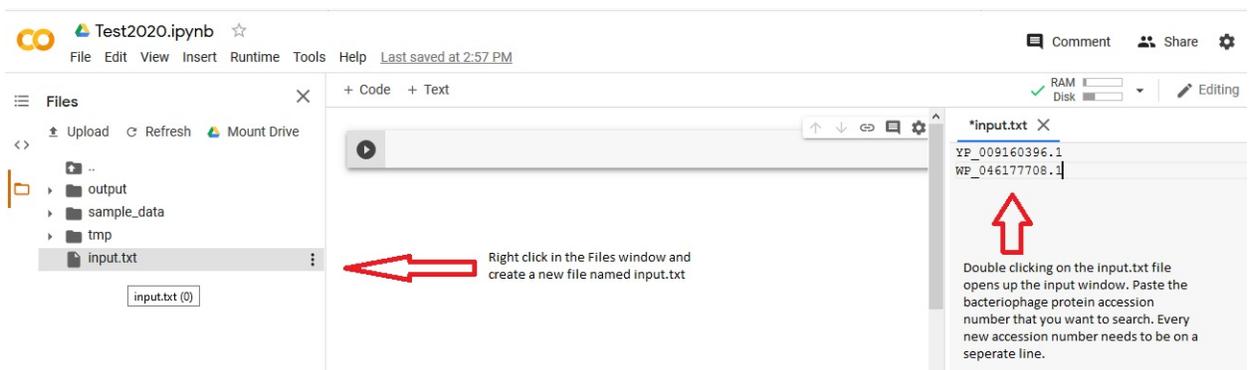
2. Rename your notebook. Then click on the File tab to open it.



3. In the File window that opens up – Right click (PC users) or CTRL + Click (Mac users) to create two new folders names – “output” and “tmp”

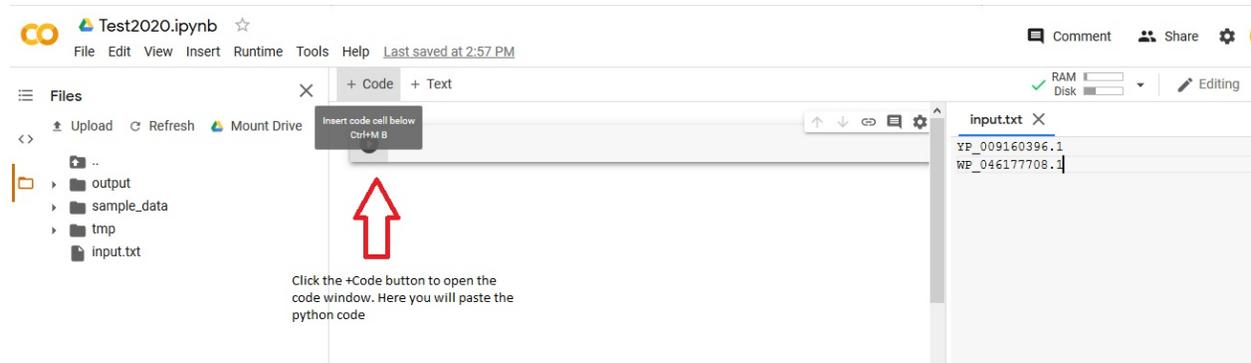


4. Right click and create a file named “input.txt”. Clicking on the file will open it up in the main window – Paste your phage protein accession numbers in the file on separate lines. Sample accession numbers can be found here: <https://github.com/genomesolver/CSPpipeline/blob/master/input.txt>

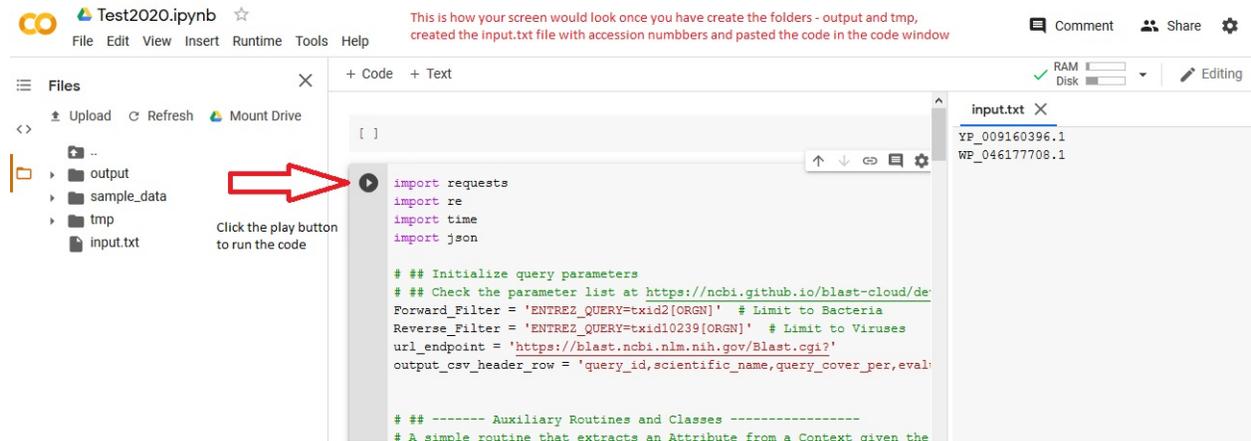


5. Click the **+Code** button in the central panel to open the code window. Here you will paste the python code. Code can be copied from here:

<https://raw.githubusercontent.com/genomesolver/CSPpipeline/master/blast.py>



6. Based on the instructions above, your collaborate window will look like the screenshot below. The final step is click on the play button to run the code



7. The output folder will contain the results of the “Forward” and “Reverse” Blast. You can download the output data in the .csv format and upload your results to the Community Science Pipeline database: [https://docs.google.com/forms/d/e/1FAIpQLSe8SmjjAdZrAKWBWlxcpmICIQTV\\_FC1cgKWgUL\\_OqWdyRkCgg/viewform](https://docs.google.com/forms/d/e/1FAIpQLSe8SmjjAdZrAKWBWlxcpmICIQTV_FC1cgKWgUL_OqWdyRkCgg/viewform)

The output folder will contain the results of the Forward and Reverse Blast. You can click on the three dots to download the file

```

Check the status via web-browser:
https://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&FORMAT=OBJECT=SearchI
Will wait and check in 60 seconds
We got some hits, will grab the json hits file
Downloaded the JSON hits file
#####
Query Cover % E_Value Accession Id Subject Name
#####
99.81 0 QCS36701 Capybara microvirus Cap3_SP_645
99.81 0 QCS36696 Capybara microvirus Cap3_SP_641
99.81 0 AXH75067 Microviridae sp.
99.81 0 AXH78085 Microviridae sp.
99.81 0 AXH77396 Microviridae sp.
99.25 0 YP_009160396 Gokushovirinae Fen7875_21
99.81 0 AVQ10172 Gokushovirinae environmental samples
99.81 0 AXH75893 Microviridae sp.
97.39 0 AXH77863 Microviridae sp.
99.81 0 AXL14535 Microviridae sp.
#####
#### TOP HIT = QCS36701 Capybara microvirus Cap3 SP 645

```

query_id	scientific_name	query_c
YP_009160396.1	Actinobacteria bacterium	98.90510
YP_009160396.1	Burkholderiaceae bacterium	98.90510
YP_009160396.1	Flavobacteriia bacterium	98.90510
YP_009160396.1	Chlamydia trachomatis	99.27007
YP_009160396.1	Gramella jeungdoensis	99.27007
YP_009160396.1	Candidatus Roizmanbacteria bacterium	99.27007
YP_009160396.1	Gammaproteobacteria bacterium	99.27007
YP_009160396.1	Phascolarctobacterium	98.72262