**OPT4e**

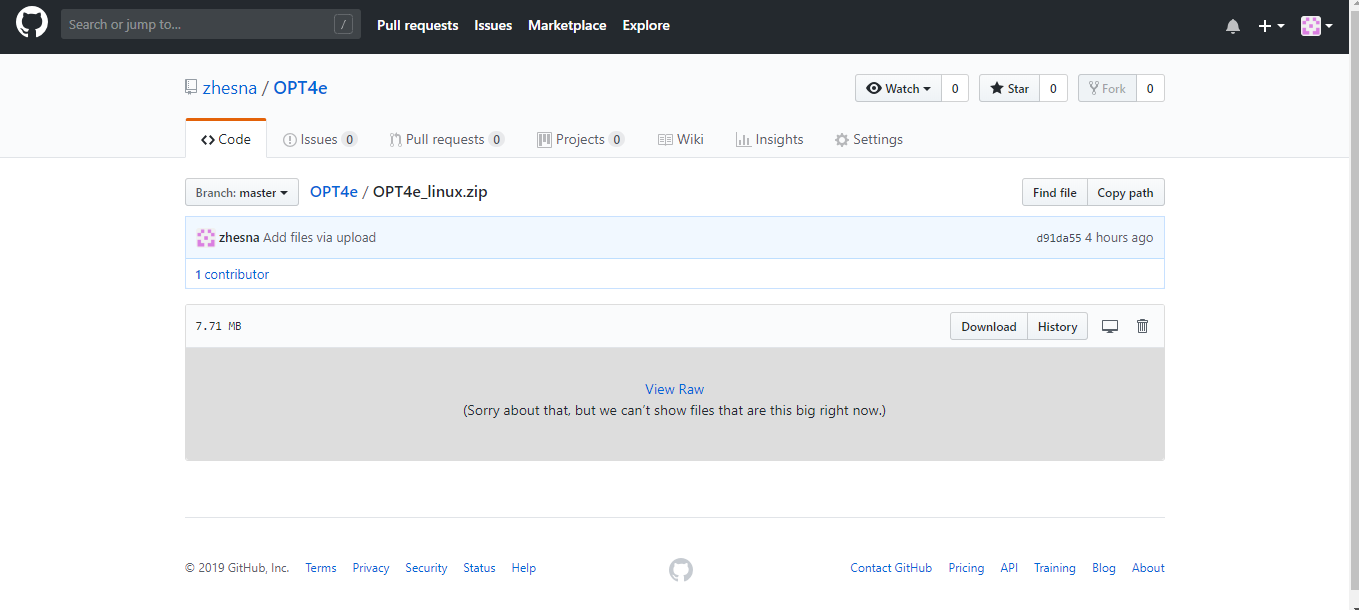
**(Optimal Prediction for T4SS Effector proteins)**

**OPT4e** is a software package written in python3 for the purpose of optimal prediction of type IV secretion system effector proteins among the input set of protein sequences. (note that this software is based on input protein sequence(s) and there is no need to provide the whole genome of a bacterial pathogen as an input.)

**Download OPT4e:**

The very first step for using OPT4e is to download its two parts.

First part is the file named “OPT4e\_linux.zip” from Github. To download it, click on the folder name, and you will be taken to a new page. Here you can click on download button, shown in this figure:



Please unzip the downloaded folder.

Second part is the folder named “FOR\_OPT4e\_linux” which is needed for using OPT4e.

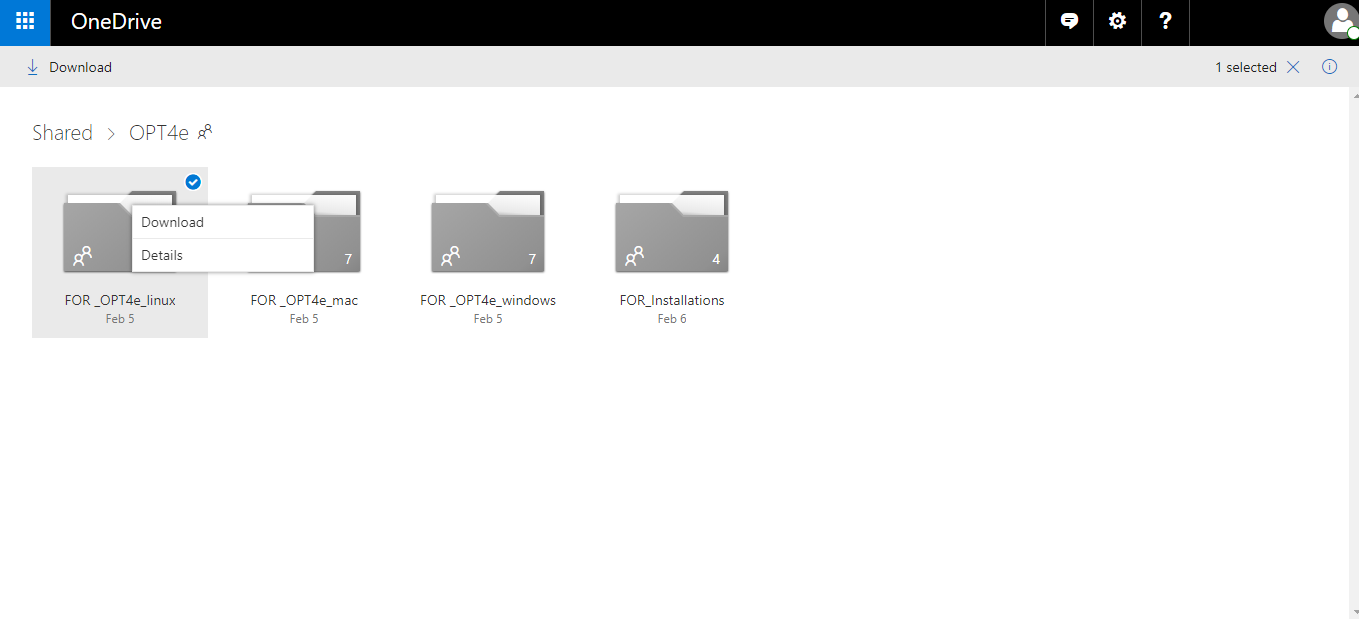
To download this folder:

**IMPRTANT:**

Using the link below, download the folder “FOR\_OPT4e\_linux” and copy all its **contents** to the **contents** of OPT4e\_windows folder, downloaded form GitHub.

<https://1drv.ms/f/s!Al5GYXWhy-wU3DgdWsdPafcQkmjh>

To download this “FOR\_OPT4e\_linux” folder, go to the above link and right-click on this folder and click download, shown in this figure:



It will download the zipped folder. Please unzip it after downloading and as mentioned, add its contents to “OPT4e\_linux” folder contents.

The rest of the manual has two parts: Installations (one-time procedure) and Running

**PART1**

**Installations:**

To use OPT4e software you need to have python and Emboss installed on your system.

**1-**It is highly recommended to install “Jupyter Notebook” which will automatically add python to your system and enables you to run the OPT4e notebook (explained in next part of manual).

To install Jupyter Notebook, please use this link (https://www.anaconda.com/download/ ) and install anaconda for python 3 which automatically adds Jupyter Notebook to your operating system. It is a very simple and fast way of adding Jupyter Notebook to all types of operating systems.

(Users who would like to install Jupyter Notebook through terminal or cmd, can see this link for explanations: <https://jupyter.org/install> )

(Also, the users who prefer to deal with python scripts, OPT4e.py script is provided to run. But it is recommended to use the Jupyter Notebook option.)

**2-**For installing Emboss:

Please go to this address.

<https://1drv.ms/f/s!Al5GYXWhy-wU3DgdWsdPafcQkmjh>

Then click on "FOR\_Installations" folder and then download the whole folder named “EMBOSS-6.6.0”.

(you can also download its tar version from this link: <ftp://emboss.open-bio.org/pub/EMBOSS/> )

Then open up a terminal and cd to this downloaded folder and install EMBOSS using these commands:

./configure –without -x

make

sudo make install

(there are some more explanations provided in their website as well if you are interested: <http://emboss.sourceforge.net/download/>)

**3-** (extra step for linux users)Please run this command in your terminal:

sudo apt-get install blast2

You are all set.

**PART2**

**Run OPT4e:**

-Before running the software, you need to provide your input protein sequence(s). The input file should has ".fa" format and should be named "input\_sequences". ***It is recommended to copy and paste your fasta format sequences in the "input\_sequences.fa" file which is provided in the OPT4e\_windows folder***.

(There are some sequences already in this file for the purpose of testing. You can simply delete them.)

-To run OPT4e, first you need to start Jupyter Notebook, which is a simple procedure. For this purpose, type “jupyter ntebook” in the terminal.

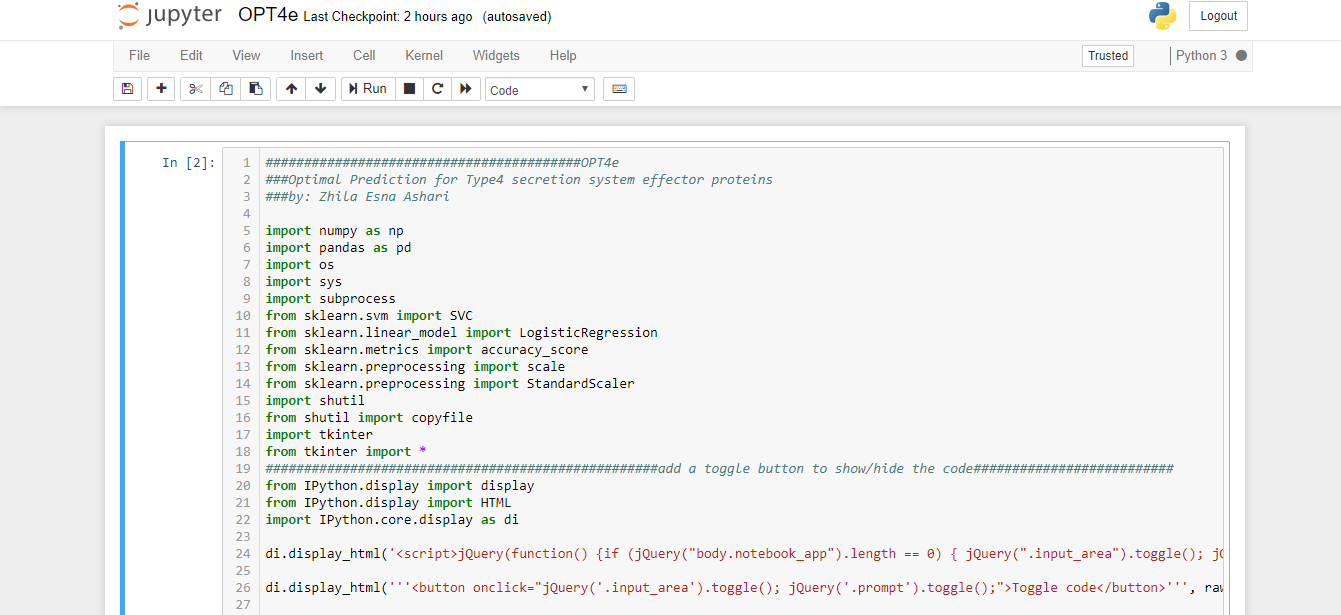
Then you should wait for some seconds, until a new page for Jupyter Notebook opens up in your default browser (for example in google chrome). This new page shows the folders in your system. Similar to this figure:



Then you should go to the place where you have saved “OPT4e\_linux” folder. For instance, if you have saved it on desktop, click on the “Desktop” shown in the above figure. Then click on the “OPT4e\_linux”. Finally, click on “OPT4e.ipynb” to open it, like this figure:

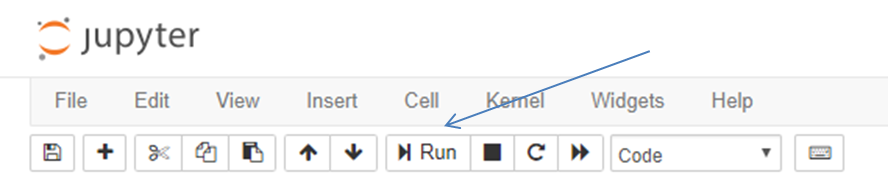


Then OPT4e.ipynb will be opened in your browser, like this figure:

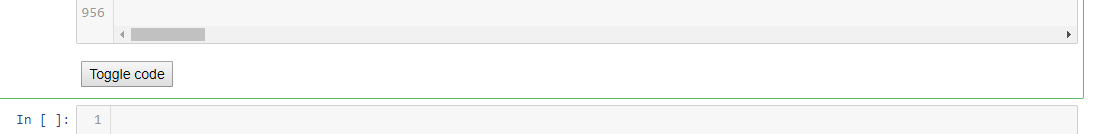


Now that you could start Jupyter Notebook, it is time to run OPT4e. First off some buttons from top bar are introduced:

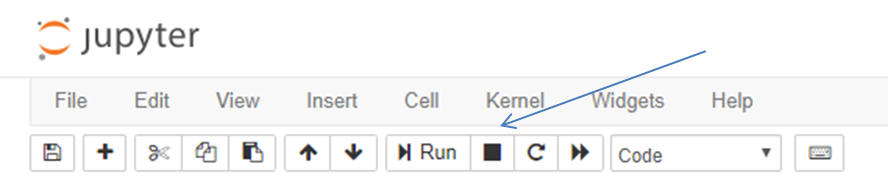
=>To start OPT4e GUI, you should just click on the RUN button from the top bar:



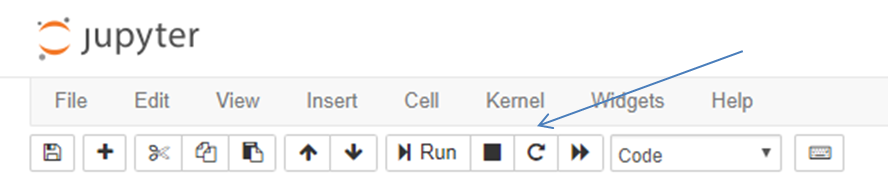
--please note that after running the OPT4e notebook, a toggle button will show up at the end of the code and you can use it to show/hide the code when using the software: (It makes it easier to work for users, like developers that want to see or not to see the code at the same time.)



=> Also, you can stop the running code whenever you want by clicking the stop button.



=> **More importantly, if an error occurs when running the code, don’t forget to click on “restart the kernel” button from the top bar before running again:**

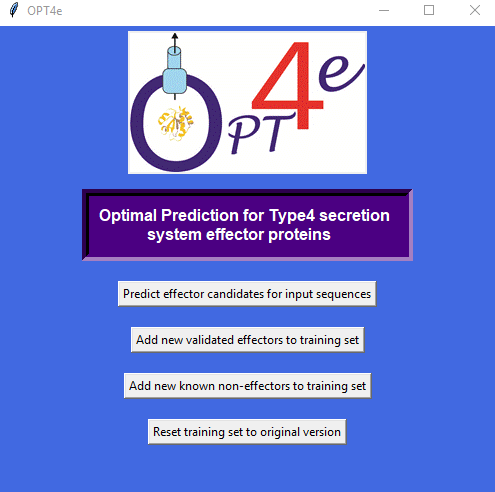


. **--***optional but recommended: If this is your first time running OPT4e, in order to make sure that you have all necessary packages in your system, you can first run "packages notebook .ipynb " . This would be just a one-time procedure. To do this, you can do the same procedure you did for opening and running OPT4e.ipynb file using run button from top bar.*

**Start OPT4e GUI:**

To start the GUI, click on the Run button form the top bar, shown before, and the GUI will be opened.

-When the GUI pops up, you have 4 buttons to choose, that are explained here:



1-"*Predict effector candidates for input sequences*": this button is used to predict if your input protein sequence(s) is a candidate T4SS effector. As mentioned, you input sequences should be in “input\_sequences.fa” file.

Please note that this step might take so long to get completed based on the number of input sequences. **Please do not close the GUI window. When the task is done, a new window will open up and let you know that it is done.**

After finishing the task, the predicted candidate effectors will be saved in 4 files (in OPT4e\_windows folder), based on their likelihood of being an effector: predicted\_effectors\_mostlikely, predicted\_effectors\_probable, predicted\_effectors\_lesslikely, predicted\_effectors\_total.

The number of predicted candidates are reported at the beginning of each file as well.

2-"*Add new validated effectors to training set*": in case you have some experimentally verified effectors and you want to add them to the software to make its future predictions better, you can use this button. As mentioned, you input sequences should be in “input\_sequences.fa” file.

**Please do not close the GUI window. When the task is done, a new window will open and let you know that it is done.**

3-"Add new validated non-effectors to training set": in case you have some known non-effectors and want to add them to the software to make its future predictions better, you can use this button. As mentioned, you input sequences should be in “input\_sequences.fa” file.

**Please do not close the GUI window. When the task is done, a new window will open and let you know that it is done.**

4-"Reset training set to original version": In case you realize that the changes you made to the training set, using the second or third button, were not correct and want to undo it, you can use this button.

**Please do not close the GUI window. When the task is done, a new window will open and let you know that it is done.**

Thank you for using OPT4e!

Zhila Esna Ashari

[z.esnaashariesfahan@wsu.edu](mailto:z.esnaashariesfahan@wsu.edu)

**Citations:**

**-**Esna Ashari Z, Brayton KA, Broschat SL*. Prediction of T4SS Effector Proteins for Anaplasma Phagocytophilum Using OPT4e*. Submitted to Frontiers in Microbiology Journal, Feb. 2019.

-Esna Ashari Z, Brayton KA, Broschat SL. *Using an optimal set of features with a machine learning-based approach to predict effector proteins for Legionella pneumophila*. PLoS ONE 2019; 14(1): e0202312. (<https://doi.org/10.1371/journal.pone.0202312>).

**-**Esna Ashari Z, Dasgupta N, Brayton KA, Broschat SL. An optimal set of features for predicting type IV secretion system effector proteins for a subset of species based on a multi-level feature selection approach. PLoS ONE 2018; 13(5): e0197041. (<https://doi.org/10.1371/journal.pone.0197041>).

-Esna Ashari Z, Brayton KA, Broschat SL. Determining Optimal Features for Predicting Type IV Secretion System Effector Proteins for Coxiella burnetii. Proceedings of 8th ACM BCB conference. 2017; 346–351. (Doi: 10.1145/3107411.3107416).