

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 set of proteins provided. (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.)

Since the installation procedure is different for each platform and it is a one-time procedure, we explain about using OPT4e in the first part of the manual. And then explain about needed installations in the second part of the manual.

First of all, using the link below, download the folder "FOR_OPT4e_windows" and copy all its contents to the main OPT4e_windows folder.

<https://1drv.ms/f/s!AI5GYXWhy-wU3HdimPf7X6XsU9Ie>

PART1

Run OPT4e:

-Before running, you need to provide your input protein sequence(s). The input file should have ".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.

-To run the software, simply run OPT4e notebook (it is highly recommended to use Jupyter Notebook version of OPT4e, but also, OPT4e.py script can be used as well.) After running the notebook, GUI will open up giving you some options to choose based on what you want to do.

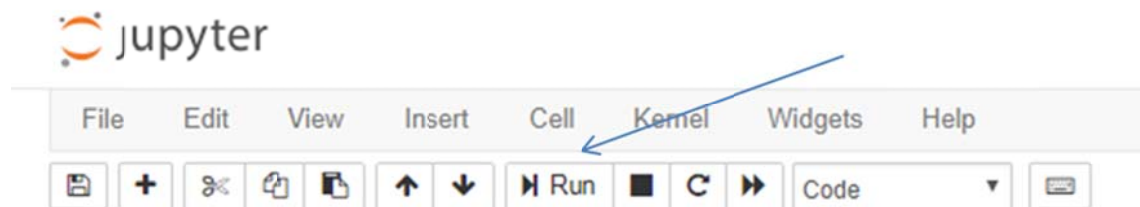
=>To run Jupyter Notebook you can search its name in your start menu in windows systems and click on its icon. See this figure:

After clicking, a terminal for jupyter notebook will be opened and then it will open a page in your browser automatically.

Then you can simply go to the folder that where you have saved OPT4e and click on “OPT4e.ipynb” to open it.

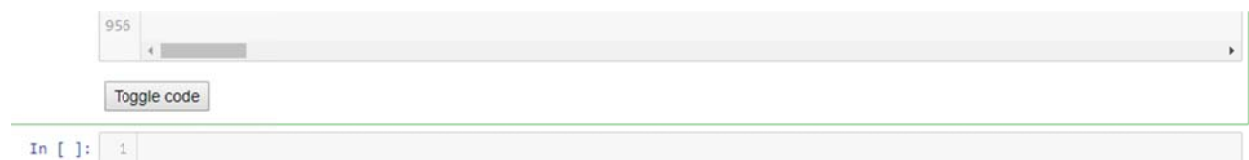
In Mac and Linux systems, you can type “jupyter ntebook” in the terminal and it will start. And the rest of the steps will be the same.

=>To run OPT4e, just click on the RUN button from the top bar:

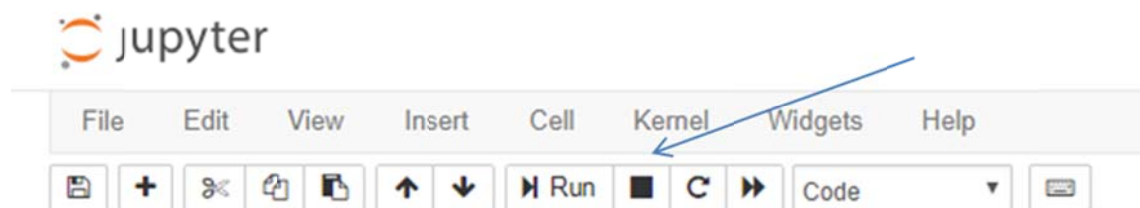


--optional but recommended: to make sure that you have all necessary packages in your system, you can first run "packages notebook" (or "packages.py" script). This would be just a one-time procedure.

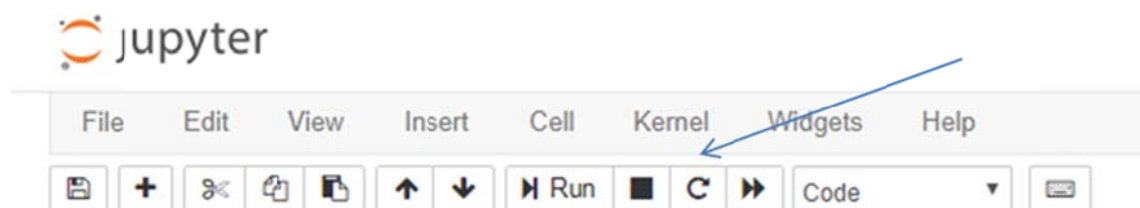
--please note that after running the OPT4e notebook, a toggle button will show up at the end of the code and you can use it to show/hide the code when using the software. It makes it easier to work for users, like developers that want to see or not to see the code at the same time.



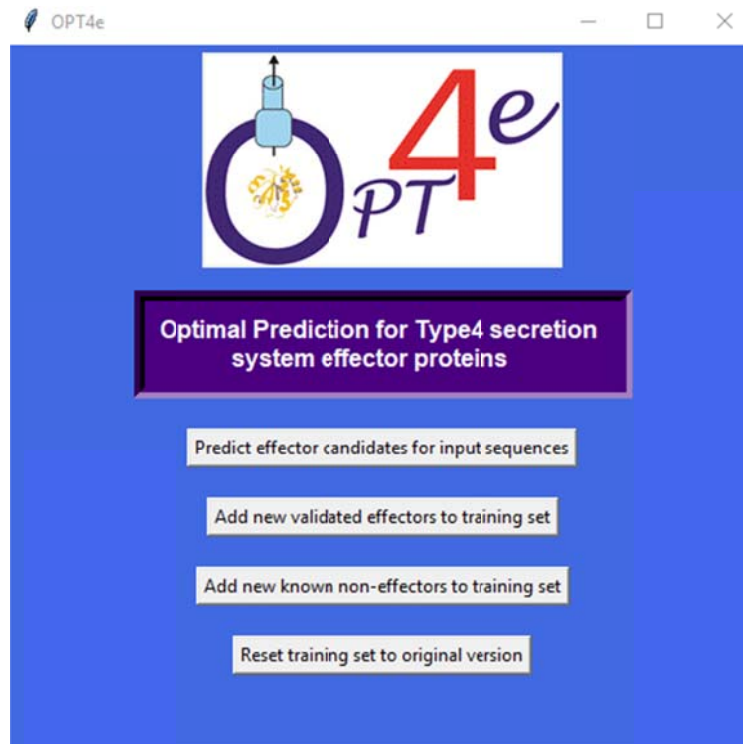
=> Also, you can stop the running code whenever you want by clicking the stop button.



=> 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:



-When the GUI pops up, you have 4 buttons to choose.



1-"Predict effector candidates for input sequences": this button is used to predict if your input protein sequence(s) is a candidate T4SS effector.

Please note that this step might take 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.

The predicted candidate effectors will be saved in 4 files (in the same folder of the software package), based on their likelihood of being an effector: predicted_effectors_mostlikely, predicted_effectors_probable, predicted_effectors_leslikely, predicted_effectors_total.

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

2-"Add new validated effectors to training set": in case you have some experimentally verified effectors and want to add them to the software to make its future predictions better, 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.

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.

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.

PART2

Installations:

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

1-For having python on your system, it is highly recommended to install Jupyter Notebook and run the OPT4e notebook. For this, you can 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 the a very simple way of adding Jupyter Notebook to all types of operating systems.

Moreover, there is another option for adding Jupyter Notebook through terminal in your system which is explained here:

<https://jupyter.org/install>

But note that for using this option, you need to have python3 installed on your system.

Although you can just have python installed on your system and run use the OPT4e.py script.

2-For installing Emboss: Go to "FOR_Installations" folder in this address:

<https://1drv.ms/f/s!AI5GYXWhy-wU3Hjzaf4DvxZKnHVu>

and download and then install "mEMBOSS-6.5.0.0-setup" which is very straightforward (you can download a folder by right clicking on it). Although, you may get a message from your system that you

need to have java installed on your system. In this case, you can use "JavaSetup8u191" from the same folder to install java.

You are all set.

Thank you for using OPT4e!

Zhila Esna Ashari

z.esnaashariesfahan@wsu.edu

Citations request:

- 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).