

OPT4e User's Guide for Windows

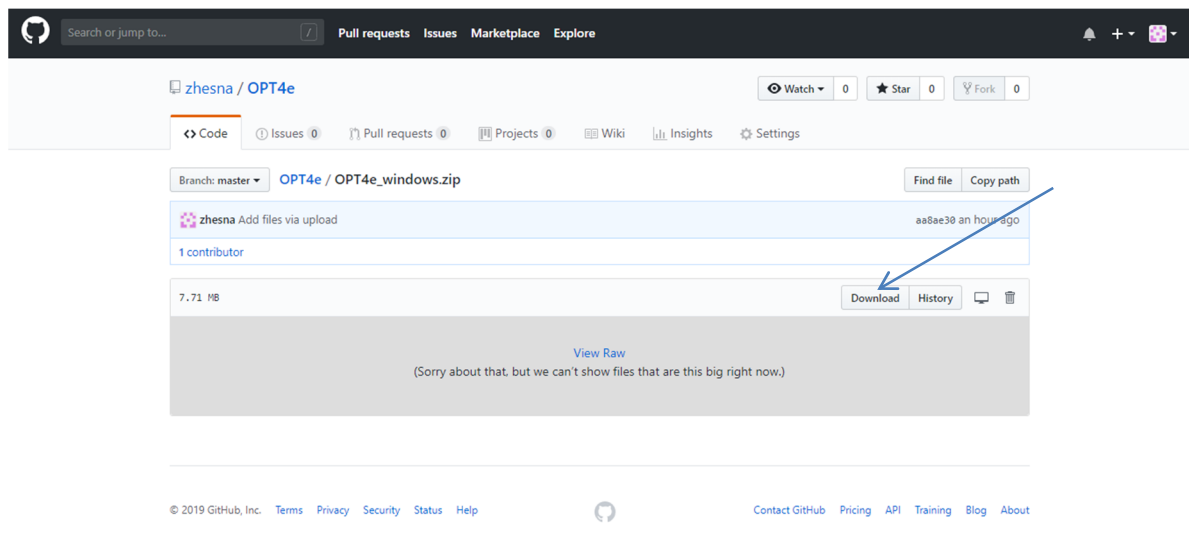
(Optimal-features Predictor for T4SS Effector proteins)

OPT4e is a software package written in Python 3 for the purpose of predicting type IV secretion system effector proteins from an 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 input data.

A. DOWNLOADING OPT4e

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

The first part is the file called “OPT4e_windows.zip” located at Github (https://github.com/zhesna/OPT4e). To download it, click on the folder name and you will be taken to a new page. Click on the Download button shown in this figure:



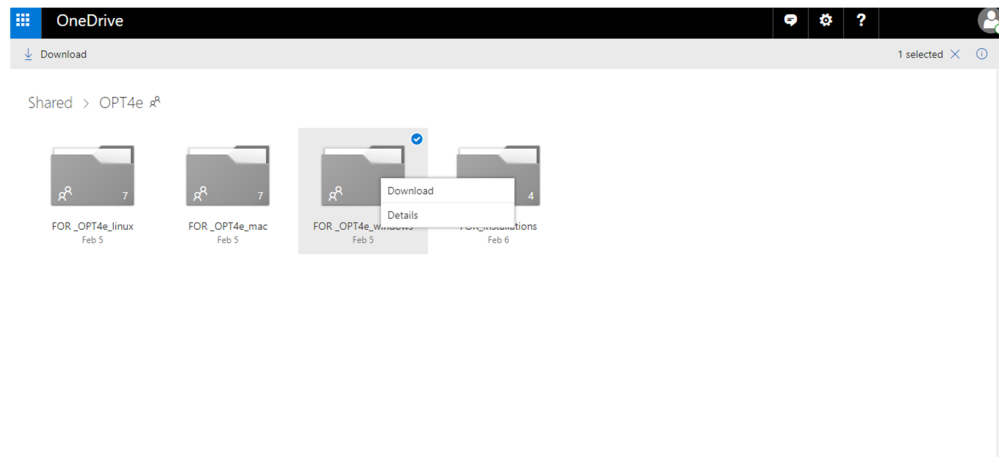
2. Next unzip the downloaded file in the location of your choice (e.g., you can create a folder called opt4e).
3. The second part of OPT4e is in the folder called “FOR_OPT4e_windows”.

IMPORTANT:

Using the link below, download the folder “FOR_OPT4e_windows” (stored on Microsoft OneDrive) and copy all its **contents** to the **contents** of your OPT4e_windows folder.

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

At this link, right-click on the folder and click on Download (see figure below):



Please unzip the folder after downloading and, as mentioned, add its contents to your “OPT4e_windows” folder contents. **Note that this folder is almost 1 GB in size so it may take a while to download.**

B. INSTALLATION OF SOFTWARE

To use the OPT4e software you must have Python and EMBOSS installed on your system. For python, You need to install Jupyter Notebook which will automatically add Python to your system.

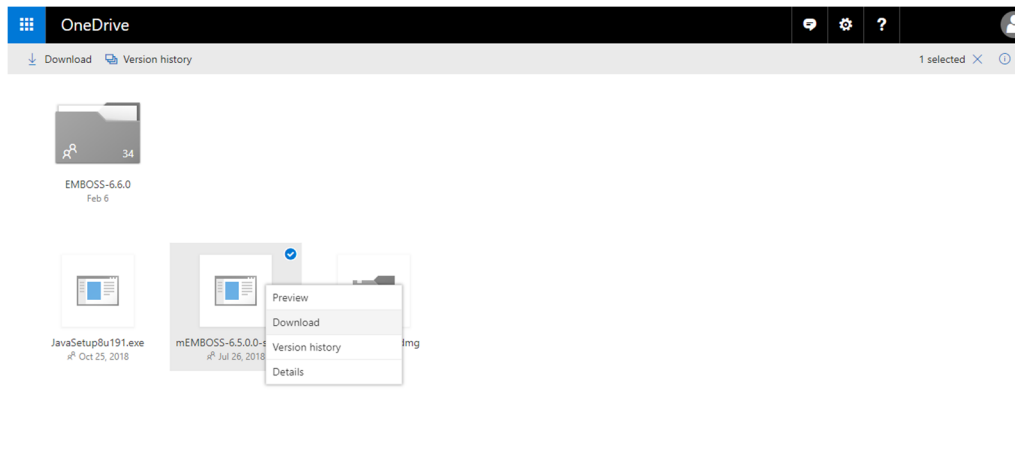
-To install Jupyter Notebook, please use this link (<https://www.anaconda.com/download/>) and install Anaconda for Python 3 which automatically adds both Jupyter Notebook and Python to your operating system. It is a very simple and fast way of adding Jupyter Notebook to all types of operating systems.

Users who would like to install Jupyter Notebook through a terminal window or cmd should go to: <https://jupyter.org/install>. Also, for users who prefer to work with Python scripts, OPT4e.py is provided. However, Jupyter Notebook is the recommended option for using OPT4e.

-For installing Emboss, go to this address:

<https://1drv.ms/f/s!AI5GYXWhy-wU3DgdWsdPafcQkmjh>

Then click on "FOR_Installations" folder. Then right click "mEMBOSS-6.5.0.0-setup" file and download it, like this figure:



Then start to install it by double clicking on the downloaded file (which is very straightforward).

-Probable step: Please note that you may get a message from your system saying that you need to have java installed on your system. Just in this case, you need to do this probable step: Please, go to this link again: <https://1drv.ms/f/s!Al5GYXWhy-wU3DgdWsdPafcQkmjh>

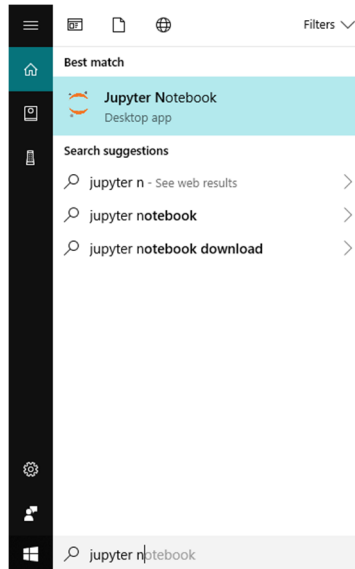
Then, go to "FOR_Installations" folder again and download "JavaSetup8u191.exe" by right clicking on it. Then start to install it by double clicking on the downloaded file. After completing its installation, go back to "mEMBOSS-6.5.0.0-setup" file and start to install it by double clicking on it.

You should now be all set to run OPT4e.

C. RUNNING OPT4e

Before running the software, you have to provide your input protein sequence(s). The input file should be in ".fa" format and should be named "input_sequences." ***It is recommended that you copy and paste your fasta format sequences into the "input_sequences.fa" file provided in the OPT4e_windows folder.*** There are some sequences already in this file for the purpose of testing, which can be deleted. **DO NOT edit any other files in the OPT4e_windows folder. Only modify the input_sequences.fa file; OPT4e will use this file differently depending on which OPT4e function you use.**

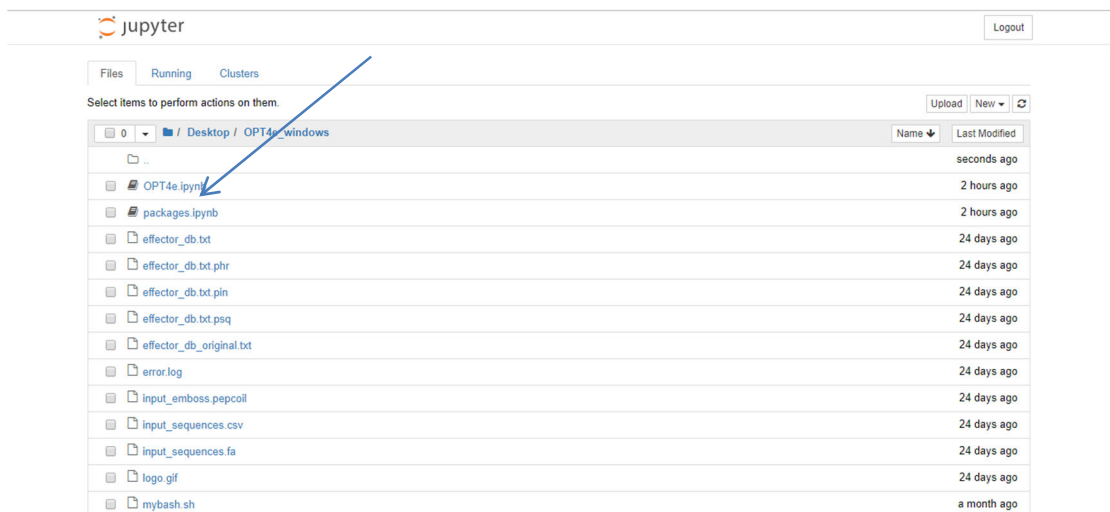
-To run OPT4e, start Jupyter Notebook, which is a simple procedure shown in pictures here. For this purpose, type "Jupyter Notebook" in the start section of you windows system and click on the orange icon for Jupyter Notebook. It is shown in this figure:



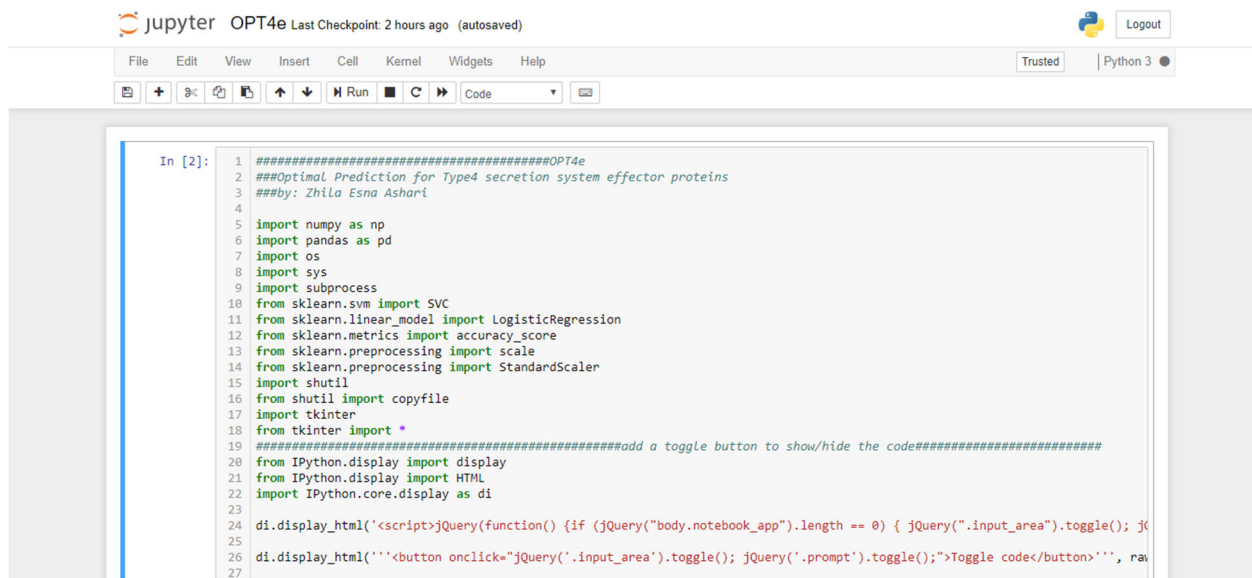
When you click on it, a Jupyter Notebook terminal will be opened. Then you should wait for some seconds, until a new page for Jupyter Notebook opens up in your default browser (e.g., in google chrome). The browser will show the folders in your system similar to the following figure:



Navigate to the place where you saved the “OPT4e_windows” folder. For instance, if you saved it on the Desktop, click on the “Desktop” shown in the above figure. Next click on the “OPT4e_windows” folder. Finally, click on “OPT4e.ipynb” to open it:



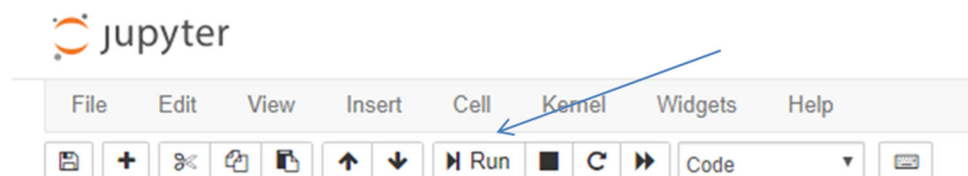
OPT4e.ipynb will open in your browser and should look similar to the following:



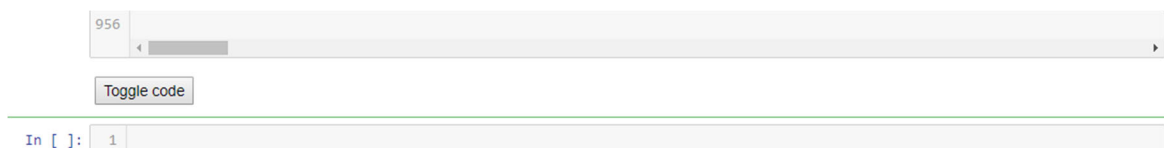
```
In [2]: 1 #####OPT4e
2 ###Optimal Prediction for Type4 secretion system effector proteins
3 ###by: Zhila Esna Ashari
4
5 import numpy as np
6 import pandas as pd
7 import os
8 import sys
9 import subprocess
10 from sklearn.svm import SVC
11 from sklearn.linear_model import LogisticRegression
12 from sklearn.metrics import accuracy_score
13 from sklearn.preprocessing import scale
14 from sklearn.preprocessing import StandardScaler
15 import shutil
16 from shutil import copyfile
17 import tkinter
18 from tkinter import *
19 #####add a toggle button to show/hide the code#####
20 from IPython.display import display
21 from IPython.display import HTML
22 import IPython.core.display as di
23
24 di.display_html('<script>jQuery(function() {if (jQuery("body.notebook_app").length == 0) { jQuery(".input_area").toggle(); j(
25
26
27 di.display_html('<<button onclick="jQuery(\'.input_area\').toggle(); jQuery(\'.prompt\').toggle();">Toggle code</button>\'', raw
```

Now that you have opened OPT4e in the Jupyter Notebook, you can run OPT4e. Look at the buttons near the top menu bar:

⇒ To start the OPT4e GUI, click on the Run button shown below:



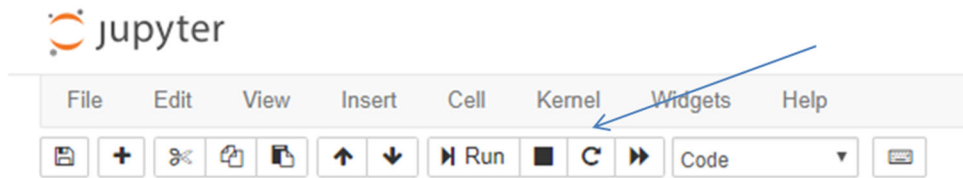
Please make sure to click inside the code cell before using the Run button. Also, note that after the OPT4e notebook has begun running, a toggle button will show up at the end of the code which you can use to show/hide the code while using the software.



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



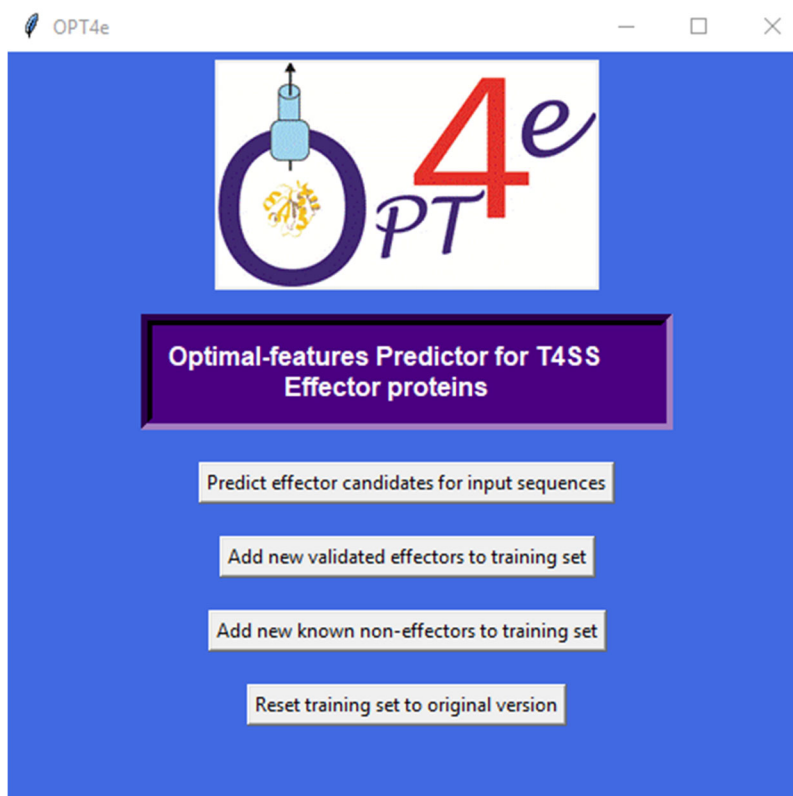
=> **Importantly, if an error occurs while running the code, click on the “restart the kernel” button near the top bar before running OPT4e again:**



Optional but recommended: In order to make sure you have all the necessary packages in your system the first time you run OPT4e, you can first run "packages.ipynb". You only have to do this once. Rather than choosing OPT4e.ipynb, choose packages.ipynb and click on the Run button. If you have all the necessary packages in your Jupyter Notebook, there will be no output. If a package or packages are missing, a message about the ones that you don't have and need to be added to your Jupyter Notebook will be displayed.

D. Using the OPT4e GUI

To start the GUI, click on the Run button near the top bar, shown before, and the GUI will open. There are four buttons on the GUI from which to choose:



- 1) *“Predict effector candidates for input sequences”*: Use to predict which input protein sequence(s) are candidate T4SS effector proteins. **Importantly, input sequences should be in the “input_sequences.fa” file.**

Please note that this step might take a considerable amount of time depending on the number of input sequences. **Please do not close the GUI window and make sure that your computer won’t go to sleep. When the task is done, a pop-up window will appear to announce this.**

After prediction has been completed, candidate effectors will be saved in four files in your 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 is reported at the beginning of each file as well. **Please make sure that you don’t have these files open while running OPT4e so that the program can re-write to them.**

- 2) *“Add new validated effectors to training set”*: Use to add newly-validated effectors to improve predictions. **Importantly, these new sequences should be added to the “input_sequences.fa” file. Do not change any other files.**

Please do not close the GUI window and make sure that your computer won’t go to sleep. When the task is done, a pop-up window will appear to announce this.

- 3) *“Add new validated non-effectors to training set”*: Use to add known non-effectors to improve predictions. **Importantly, these new sequences should be added to the “input_sequences.fa” file. Do not change any other files.**

Please do not close the GUI window and make sure that your computer won’t go to sleep. When the task is done, a pop-up window will appear to announce this.

- 4) *“Reset training set to original version”*: Use to reverse changes you made to the training set in 2) and 3) in case of error or change of mind.

Please do not close the GUI window. When the task is done, a pop-up window will appear to announce this.

Thank you for using OPT4e!

Zhila Esna Ashari

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. (<https://doi.org/10.1145/3107411.3107416>).