Manual of PlasGUN Release Version 1.0

Table of Contents

Any question, please do not hesitate to contact me: fangzc@pku.edu.cn

1. Operating system

Linux (PlasGUN has been tested on Ubuntu 16.04)

2. Requirements

●Python 2.7.12 (https://www.python.org/)
●Python packages:
  ○numpy 1.13.1(http://www.numpy.org/)
  ○h5py 2.6.0(http://www.h5py.org/)
●TensorFlow 1.4.1 (https://www.tensorflow.org/)
●Keras 2.0.8 (https://keras.io/)
●MATLAB Component Runtime (MCR) R2018a
(https://www.mathworks.com/products/compiler/matlab-runtime.html)

or

MATLAB R2018a (https://www.mathworks.com/products/matlab.html)

Note:
1. For compatibility, we recommend installing the tools with the similar version as described above.
2. If GPU is available in your machine, we recommend installing a GPU version of the TensorFlow to speed up the program.
3. PlasGUN can be run with either an executable file or a MATLAB script. If you run PPR-Meta through the executable file, you need to install the MCR while MATLAB is not necessary. If you run PPR-Meta through the MATLAB script, MATLAB is required.

3. Preparation

Please install numpy, h5py, TensorFlow, Keras, MCR (or MATLAB) according to their manuals.

The numpy, h5py, TensorFlow, and Keras are python packages, which can be installed with "pip". If "pip" is not already installed in your machine, use the command "sudo apt-get install python-pip python-dev" to install "pip". Here are example commands of installing the above python packages using "pip".

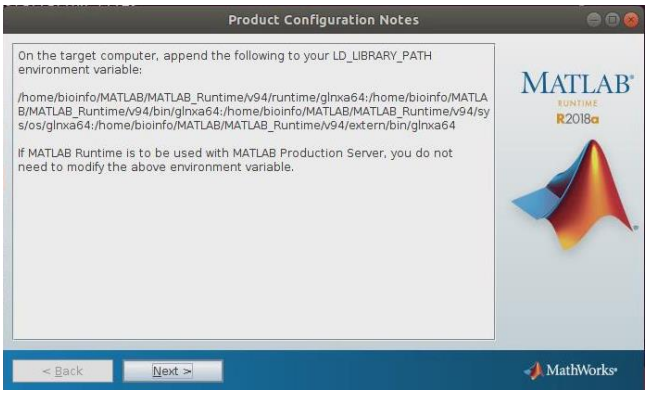
pip install numpy
pip install h5py
pip install tensorflow==1.4.1 #CPU version
pip install tensorflow-gpu==1.4.1 #GPU version
pip install keras==2.0.8

If you are going to install a GPU version of the TensorFlow, specified NVIDIA software should be installed. See https://www.tensorflow.org/install/gpu to know whether your machine can install TensorFlow with GPU support.

When running PPR-Meta through the executable file, MCR should be installed. See https://www.mathworks.com/help/compiler/install-the-matlab-runtime.html to install MCR. On the target computer, please append the following to your LD_LIBRARY_PATH environment variable according to the tips of MCR:

<MCR_installation_folder>/v94/runtime/glnxa64
<MCR_installation_folder>/v94/bin/glnxa64
<MCR_installation_folder>/v94/sys/os/glnxa64
<MCR_installation_folder>/v94/extern/bin/glnxa64

A screenshot of the tips when installing MCR is shown below:



When running PPR-Meta through the MATLAB script, please see https://www.mathworks.com/support/ to install the MATLAB.

4. Usage

To run PlasGUN, please download PlasGUN package as a zipped file, and then unpack the zipped file and change directory to PlasGUN:

wget http://cqb.pku.edu.cn/ZhuLab/PlasGUN/PlasGUN_v_1_0.zip
unzip PlasGUN_v_1_0.zip
cd PlasGUN_v_1_0

4.1. Run by executable file (in command line)

In this form, please simply executes the command:

./PlasGUN <input_file_folder>/input_file.fna <output_file_folder>/output_file_prefix

The input file must be in fasta format. Users do not need to specify the extension name of the output file. You can use the file “example.fna” which contains 100 sequences in the program folder to test the PlasGUN by simply executing the command:

./PlasGUN example.fna result

4.2. Run by MATLAB script (in MATLAB GUI)

In this form, please execute the following command directly in the MATLAB command window.

PlasGUN('<input_file_folder>/input_file.fna', '<output_file_folder>/output_file_prefix')

For example, if you want to predict genes in example.fna, please execute:

PlasGUN('example.fna', 'result')

Remember to set the working path of MATLAB to the program folder before running the program.

4.3. Run with specified threshold

For each candidate ORF, PlasGUN calculates a score (between 0 to 1), representing the probability that the ORF belongs to coding ORF or non-coding ORF. By default, the ORF with a score higher than the default threshold is predicted as coding ORF. The selection of the default threshold is shown in the publication. Users can also specify a threshold different from the default threshold. In general, with a higher threshold, the sensitivity [Sn=TP/(TP+FN)] will be lower while the specificity [Sp=TP/(TP+FP)] will be higher. For example, if you want to take 0.7 as the threshold, please execute:

./PlasGUN example.fna result 0.7 (Run by executable file)

or

PlasGUN('example.fna', 'result','0.7') (Run by MATLAB script)

The performance of PlasGUN under different thresholds are shown in Figure S2 of the publication.

Note: When running PlasGUN, you can ignore the warning about the information of the CPU/GPU, as shown in the screenshot below.

```
2018-08-02 07:59:08.760305: I tensorflow/core/platform/cpu_feature_guard.cc:137] Your CPU supports instructions that this TensorFlow binary was not compiled to use:
 SSE4.1 SSE4.2 AVX AVX2 FMA
2018-08-02 07:59:09.081744: I tensorflow/stream_executor/cuda/cuda_gpu_executor.cc:892] successful NUMA node read from SysFS had negative value (-1), but there must
 be at least one NUMA node, so returning NUMA node zero
2018-08-02 07:59:09.082159: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1030] Found device 0 with properties:
name: GeForce GTX 1060 6GB major: 6 minor: 1 memoryClockRate(GHz): 1.835
pciBusID: 0000:01:00.0
totalMemory: 5.92GiB freeMemory: 5.62GiB
2018-08-02 07:59:09.082179: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1120] Creating TensorFlow device (/device:GPU:0) -> (device: 0, name: GeForce GTX 106
0 6GB, pci bus id: 0000:01:00.0, compute capability: 6.1)
```

5. Output

PlasGUN outputs a tabular file. The file contains the header of each sequence (starts with ">"), the coordinates of the predicted ORF, and the strand in which the ORF is located. If the ORF is located in the '+' strand, the first coordinate is the 5' end of the ORF and the second coordinate is the 3' end of the ORF; if the ORF is located in the '-' strand, the first coordinate is the 3' end of the ORF and the second coordinate is the 5' end of the ORF. The coordinates outside the bracket correspond to the longest ORF. The coordinate in the bracket of the 5' end is the translation initiation site selected by MetaTISA.

	A	B	C
1	>read1		
2	3(3)	341	+
3	>read2		
4	2(2)	352	+
5	>read3		
6	3	173(173)	-
7	>read5		
8	1	108(108)	-
9	>read7		
10	1(1)	102	+