

CDS Filtering Program

User Manual



Yoann Pageaud.

KnopLab, ZMBH – Universität Heidelberg

May 13th, 2016

Presentation of CDS Filtering

CDS Filtering is a little python program allowing you to filter your predicted ORFs, CDSs, RNA, or any other type of DNA or RNA sequences.

It takes in input .txt, .fasta or .cdsexons files. You can also try with other format file, but there is no guarantee on the results.

Warning: the file has to contain 1 sequence per line, and 1 line for 1 sequence.

3 Filters are available in the program:

- The first filter allow you to remove the sequences present in several copies in your file to keep only one version.
- The second filter allow you to set a minimum threshold on the length of your sequence. All the sequences above this threshold will be kept, and all the sequence under will be deleted in the output file.
- The third filter allow you to set a maximum threshold on the length of your sequence. All the sequences under this threshold will be kept, and all the sequence above will be deleted in the output file.

When the filtering is done, you will get 2 output files:

- A summary of the filtering of your sequences which will contain all information about the parameters used and the statistical results of your filtering.
- An output file which will contain all your sequences conserved after the filtering.

If you are satisfied with the results of your filtering it is highly recommended to mention the program in your article references:

Pageaud, Yoann. “CDS Filtering Program”. KnopLab, ZMBH - Universität Heidelberg. May 13, 2016.

Setup

Prerequisite before using CDS Filtering

Install Python version 2.7.11 (or later version):

- **Under Windows:**
Download link: <https://www.python.org/downloads/release/python-2711/>
- **Under Linux:**
Python 2.7 should already be installed. If it is not the case:
 1. Open the Linux Terminal.
 2. Type:

```
sudo apt-get install python2.7
```

Launching under Windows

To open the terminal Windows:

1. Type: key Windows + R.
2. In the text field on the right of « Open: », type « cmd », then clic on « OK ».

In your terminal :

1. The terminal opens in a new window, a command line of type « C:\> » appears. This command line determine the repertory in which you are located to launch you program.

- To change drive, type just after the command line the name of the drive on which you would like to work on.

Example:

```
C:\>E:
```

- To change repertory in a drive : in the windows explorer, open the folder where your program is, type just after the command line the path to the repertory preceded by « cd ».

Example:

```
E:\>cd Folder\SecondFolder\MyProgramFolder
```

2. Once the drive and the repertory that contain the program are chosen, run the GOLIAT program in Python with the example command line below.

Example:

```
E:\Folder\SecondFolder\MyProgramFolder>py MyProgram.py
```

To Launch the CDS Filtering program:

```
C:\>E:
E:\>cd Folder\SecondFolder\MyProgramFolder
E:\Folder\SecondFolder\MyProgramFolder>py MyProgram.py
```

The CDS Filtering program will start. Follow the instructions.

Launching under Linux

1. Launch the Linux Terminal

2. Check your working directory:

Example:

```
user@users-desktop:~$ pwd
```

If you just start the terminal your working directory should be:

```
/home/user
```

3. Check the files and folder contained in your personal folder with the command ls:

Example:

```
user@users-desktop:~$ ls
```

4. Set the working directory to where the CDS Filtering program is saved with the command cd (if you just download the program the CDS Filtering folder is in your Downloads folder :

Example:

```
user@users-desktop:~$ cd Downloads/CDS Filtering
```

5. Launch the CDS Filtering program:

Example:

```
user@users-desktop:~$ python CDSFiltering.py
```

The CDS Filtering program will start. Follow the instructions.