

Bioinformatics Docker Images Project

Phenotypic Evolution Group - IBMC/i3S



Feel free to contact us:
pegi3sdocker@gmail.com



<https://pegi3s.github.io/dockerfiles/>

Pedro Ferreira^{1,2} | Hugo López-Fernández^{3,4,5} | Pedro Duque^{1,2} | Cristina P. Vieira^{1,2} | Florentino Fdez-Riverola^{3,4,5} | Miguel Reboiro-Jato^{3,4,5} | Jorge Vieira^{1,2} |

¹Instituto de Biologia Molecular e Celular (IBMC), Porto, Portugal | ²Instituto de Investigação e Inovação em Saúde (i3S), Universidade do Porto, Porto, Portugal | ³ESEI – Escola Superior de Engenharia Informática, Universidad de Vigo, Ourense, Spain | ⁴Centro de Investigaciones Biomédicas, Vigo, Spain | ⁵SING Research Group, Spain |

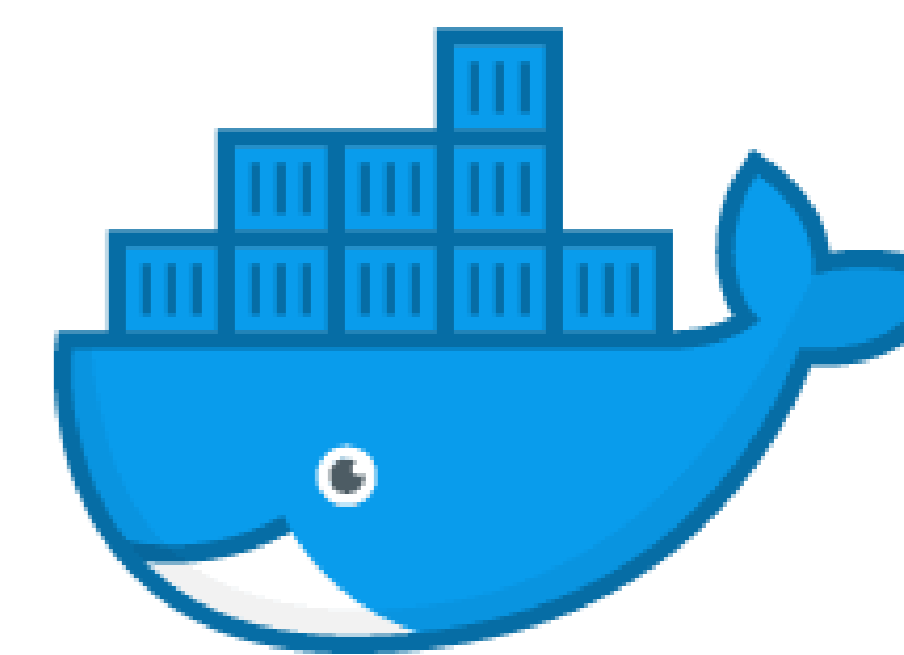
Motivation

Nowadays bioinformatics is one of the most important areas in modern biology. Nevertheless, **the use of scientific software is not always easy for researchers without an informatics background.** This is mainly due to:

- 1) Availability of the software for **Linux operating systems only**, since most researchers working on the life sciences use Windows or Macintosh operating systems;
- 2) The **complexity of the installation**, since scientific software often have many dependencies and some may be non-trivial to install;
- 3) The lack of knowledge regarding alternatives that mitigate the previous drawbacks, and how to use them.

Docker Images Advantages

- 1) Easy to use since it only requires the installation of Docker;
- 2) Portable between computers and immutable;
- 3) Can be deposited in public databases such as **Docker Hub**;
- 4) Can be downloaded when needed and erased when no longer needed;
- 5) Presents a user friendly interface – **Portainer.io** – that allows users to build and manage Docker environments easily;
- 6) The installed scientific software is ready to run;
- 7) Software invoked through the command line can be directly run in **Linux and Windows** platforms;
- 8) Docker images can be used in pipelines;
- 9) All Docker images can be run using a **VirtualBox** Ubuntu image with Docker previously installed.



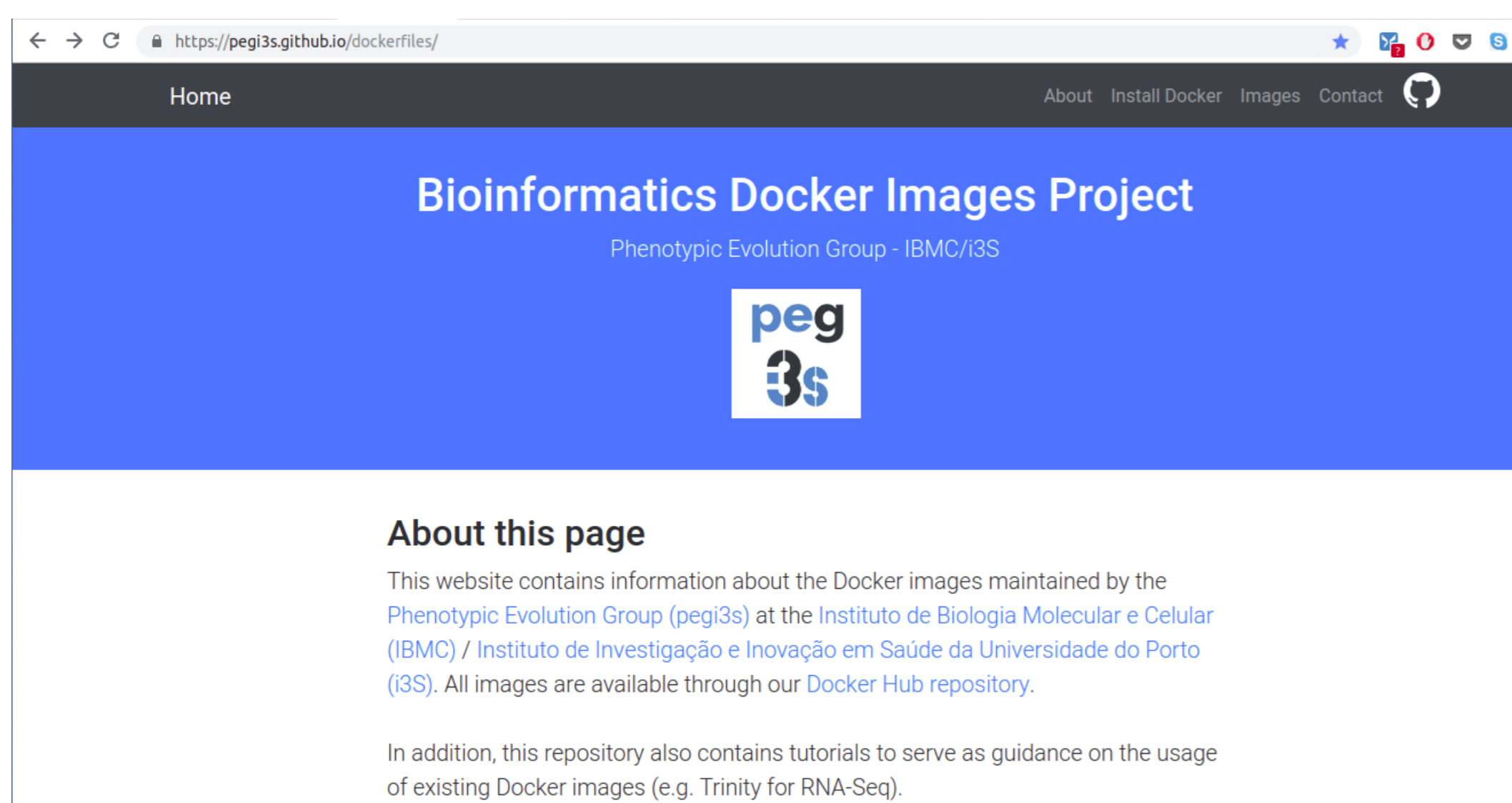
docker



1098 pulls

Website

<https://pegi3s.github.io/dockerfiles/>



Install Docker

In order to use the Docker images maintained by **pegi3s**, users need to install Docker first.

Linux

To install Docker in Linux, please follow the steps provided [here](#) (additional information can be found [here](#)).

Windows

To install Docker in Windows, please download the official installer [Docker for Windows Installer.exe](#). If your system does not meet the requirements to run Docker for Windows (available [here](#)), please install [Docker Toolbox](#) instead.

VirtualBox

As an alternative and for practical purposes, we provide an Ubuntu 18.04 LTS (Linux) virtual image with 11.8 GB, with Docker and [Portainer](#) already installed. Please note that this virtual image is limited to 100 GB. In order to run it, first download VirtualBox [here](#). After installing VirtualBox, users just need to import the OVA file [here](#) provided. To import an existing virtual machine into VirtualBox, please follow the steps described [here](#). The administrator's user and password is **pegi3s**. The username and password for [Portainer](#) is **pegi3sdocker**.

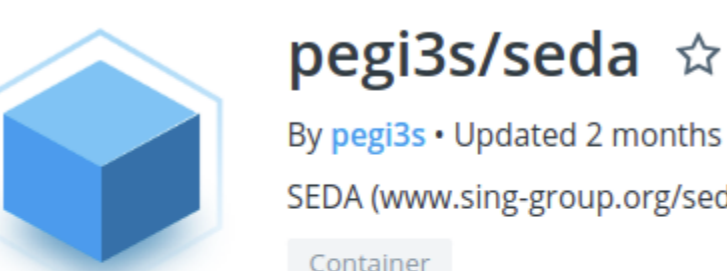
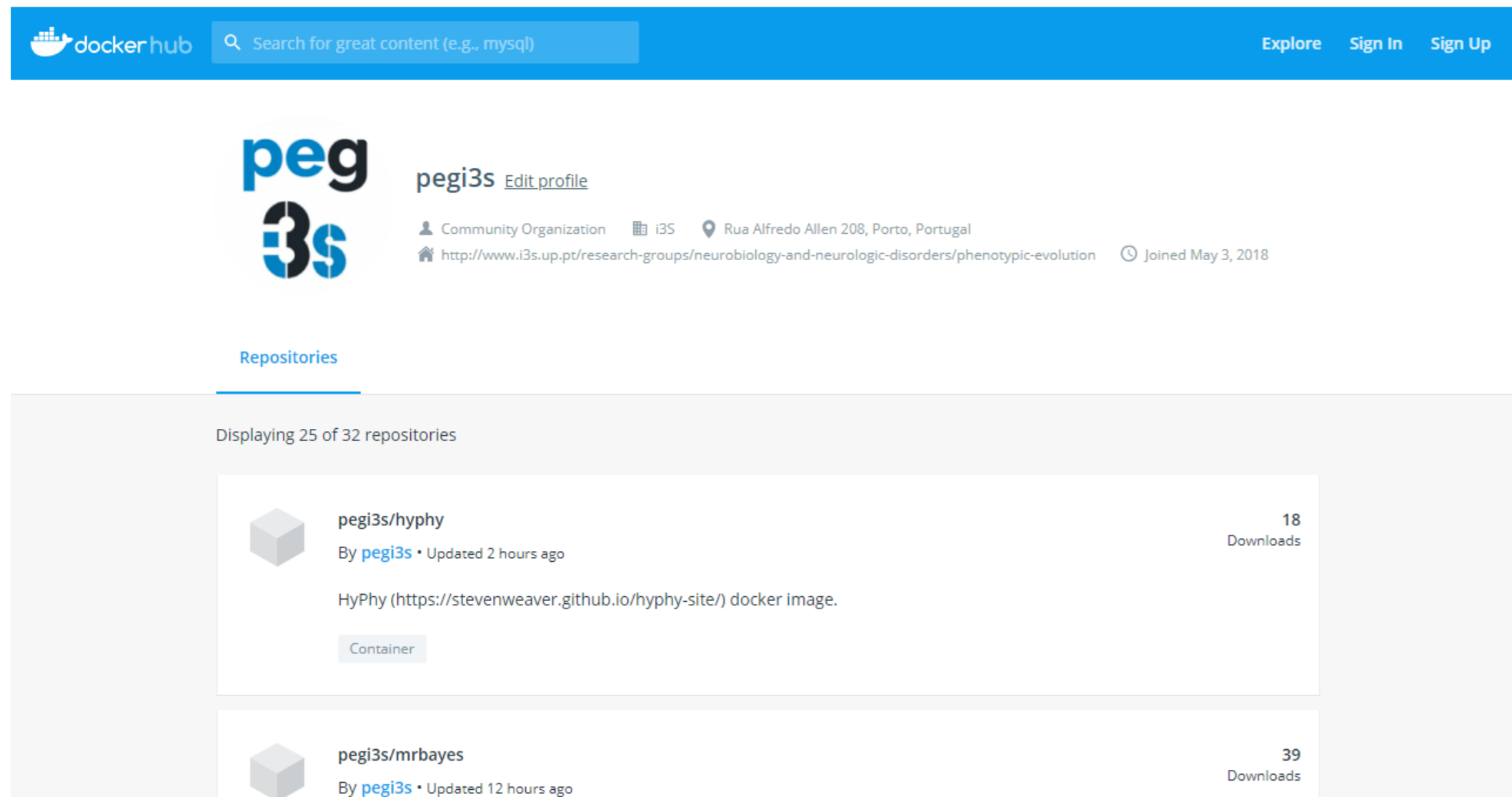
Images

List of available images:

- **abyss** [doc] - de novo sequence assembler
- **adops** [doc] - Phylogenetics and detection of positively selected sites
- **afterqc** [doc] - Sequence read quality assessment
- **alter** [doc] - File conversion
- **autoaugustus** [doc] - Gene annotation
- **bdbm** [doc] - Sequence database manager
- **bedtools** [doc] - General utilities
- **blast** [doc] - Sequence alignment
- **clustalomega** [doc] - Sequence alignment
- **coral** [doc] - Sequence read correction
- **cutadapt** [doc] - Sequence read trimming
- **edena** [doc] - de novo sequence assembler
- **emboss** [doc] - General utilities
- **fastqc** [doc] - Sequence read quality assessment
- **fasttree** [doc] - Phylogenetics inferences
- **flash** [doc] - Sequence read overlap
- **hyphy** [doc] - Phylogenetics inferences
- **mauve** [doc] - Genome alignment
- **mrBayes** [doc] - Phylogenetics inferences
- **omegamap** [doc] - Detection of positively selected sites
- **phipack** [doc] - Recombination tests
- **popart** [doc] - Population genetics software
- **raxml** [doc] - Phylogenetics inferences
- **sapp** [doc] - Genomic tools
- **seda** [doc] - Sequence dataset builder

Clear instructions on how to use!!

Docker Hub



Using the SEDA image in Linux

Please note that the following instructions must be executed in Linux environments only.

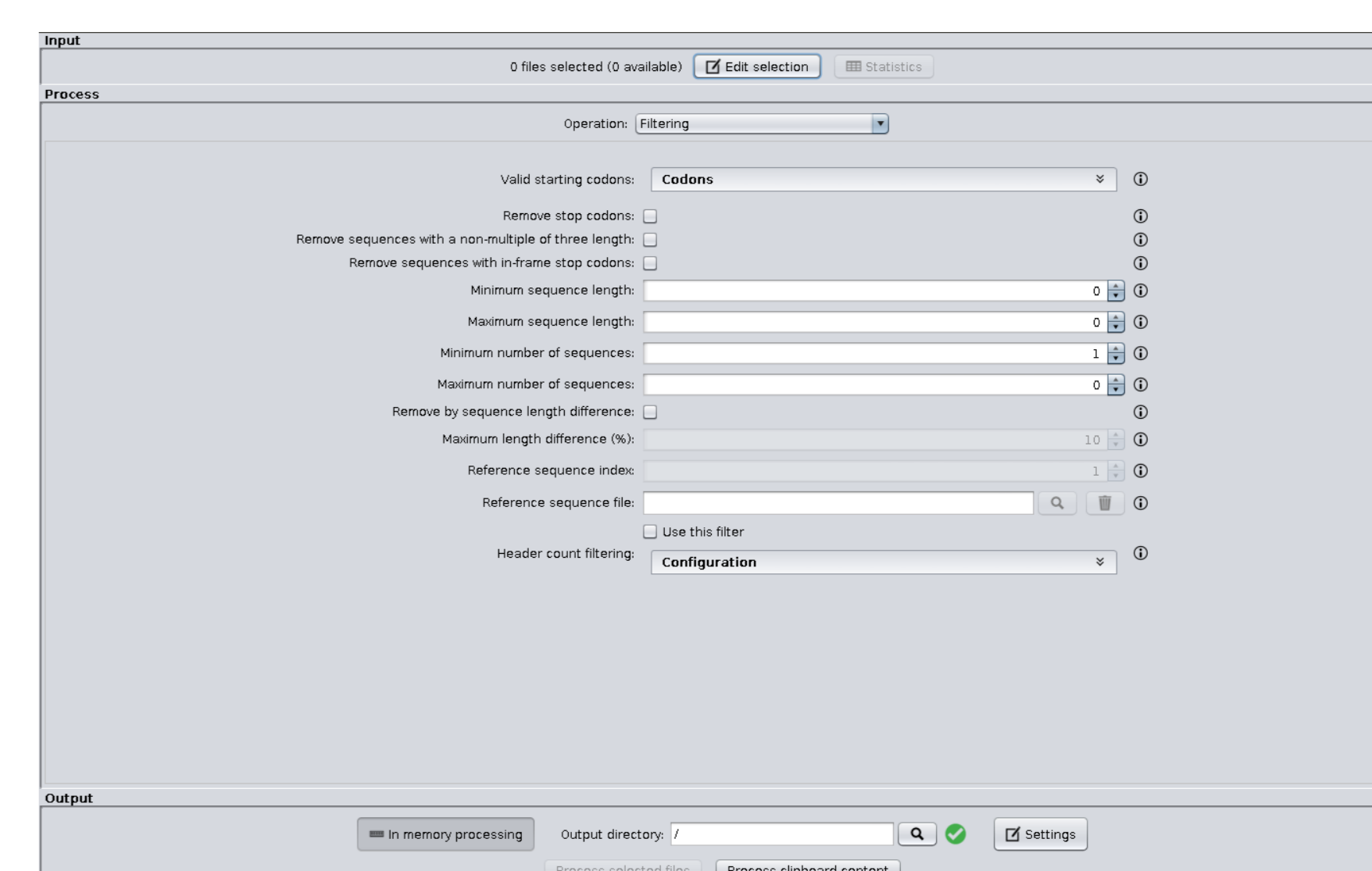
You should adapt and run the following command: `docker run --rm -it -e USERID=UID -e USER=USER -- DISPLAY=DISPLAY -v /var/lib/docker:/tmp/.X11-unix:tmp/.X11-unix -v /home/username/.Xauthority:/home/develop/.Xauthority -v /your/data/dir:/data pega3s/seda`

If the above command fails, try running `sh -c 'xhost +'` first. In this command, you should replace:

- `/your/data/dir` : point to the directory that you want to have available at `SEDA`.

Running this command opens the SEDA Graphical User Interface. Your data directory will be available through the file browser at `/data`.

SEDA Docker Image (GUI)



HyPhy Docker Image (Command Line Ubuntu 16.04 LTS)

```
root@pedro-ORTEGE-R930:/home/pedro/Desktop/IBMC/data/hyphy/2.3.14# docker run --rm -it -v /home/pedro/Desktop/IBMC/data/hyphy/2.3.14/INPUT:/data pegi3s/hyphy
HYPHY
-----
/ HYPHY 2.3.14.20190213beta(MP) for Linux on x86_64 \
***** TYPES OF STANDARD ANALYSES *****

(1) Selection Analyses
(2) Evolutionary Hypothesis Testing
(3) Relative evolutionary rate inference
(4) Coevolutionary analysis
(5) Basic Analyses
(6) Codon Selection Analyses
(7) Compartmentalization
(8) Data File Tools
(9) Miscellaneous
(10) Model Comparison
(11) Kernel Analysis Tools
(12) Molecular clock
(13) Phylogeny Reconstruction
(14) Positive Selection
(15) Recombination
(16) Selection/Recombination
(17) Relative Rate
(18) Relative Ratio
(19) Substitution Rates

Please select type of analyses you want to list (or press ENTER to process custom batch file):1
```

Conclusions

- We provide **Docker images for more than 30 scientific softwares**, commonly used in genomics, transcriptomics, sequence analysis, and phylogenetics, and **new images are constantly being added**;
- **Clear instructions on how to use the images** and when appropriate providing test cases. Links to the scientific software developer's manuals;
- We provide a **VirtualBox Ubuntu image with Docker installed** for those users not familiar with Linux operating systems;
- **Main difference to BioContainers**: detailed information that is given allowing researchers without a background in informatics to easily use these Docker images;
- **More than 1K Docker images already downloaded!**

INSTITUTO DE INVESTIGAÇÃO E INOVAÇÃO EM SAÚDE UNIVERSIDADE DO PORTO

Rua Alfredo Allen, 208
4200-135 Porto
Portugal

+351 220 408 800

www.i3s.up.pt