EDGE COVID-19

Los Alamos National Laboratory

About

EDGE COVID-19 is a bioinformatics platform that provides standardized workflows for genome 'assembly' and preliminary analysis of Illumina or Nanopore data for SARS-CoV-2 genomes. The basic workflow includes data quality control and filtering, alignment of reads to a reference genome (NC_045512.2), creation of a consensus genome sequence based on the read alignments, and a preliminary Single Nucleotide Polymorphism analysis.

The platform contains read alignment tools that accommodate Illumina or Nanopore data, including those generated using the SARS-CoV-2 ARTIC network sequencing workflows (https://artic.network/ncov-2019). It consists of a user-friendly GUI that drives a series of best-practice, open source bioinformatics software to aid in the reconstruction of complete genomes of SARS-CoV-2 from whole genome sequencing projects.

This light-weight version is available as a Docker container, able to run on any local hardware infrastructure or in the cloud, and is based on the original EDGE Bioinformatics platform, v1.1. At this moment, no reference sequence database is required, and it should be able to run on a laptop.

For users who want to use read- or assembly-based taxonomy classification tools to understand what organisms may be present within samples, we refer you to the original EDGE Bioinformatics platform which harbors this workflow but requires a number of large databases to enable such a search. In initial tests, we recover Bat Coronavirus XXX as the genome in the databases as the closest near-neighbor to the SARS-CoV-2 genome sequence, but are actively working on generating new reference indexes to also allow hits to recent SARS-CoV-2 genomes.

Users can input/upload Illumina or Nanopore sequencing fastq files (and/or download from NCBI SRA), perform *de novo* assembly, map reads to a reference genome to get the consensus SARS-CoV-2 sequence from the sequencing data. Users can also compare the assembled contigs to a reference sequence (default to NCBI refseq sequence [NC_045512.2](https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2/) and see the differences in both tabular and graphic form.

Requirements

1. Docker Engine version 19.03.2 or greater
2. The image size is around 11.8GB. This image is built on top of the official Ubuntu 18.04.3 LTS Base Image.
3. Recommended minimum computational resource : 8G memory, 4 CPUs, 20GB for the image.

How to install this image? A step by step guide:

Step 1: Install and run Docker

This step can be skipped if you have docker installed and opened in your system. If a docker instance is running in a MacOSX, dockers's icon (🐳) will show up at the top bar of your screen.

If you do not have docker installed, See https://www.docker.com/products/docker-desktop to download and install a copy.

Step 2: Obtain the docker image

On a MacOSX, open Terminal and `cd` into the directory where you want to install the image. If you want to create a new folder, then first create that folder, and then pull the docker image:

```
$ mkdir EDGE-COVID19
 $ cd EDGE-COVID19
  $ docker pull bioedge/edge_ncov
```

This can take anywhere from 10-30 minutes depending on your internet speed.

Step 3: Setup necessary databases and folders

Download MySQL database for User Management. You can either directly download from Terminal using `wget` or `curl`, or just click the [link](https://edge-dl.lanl.gov/EDGE/docker/edge_ubuntu_init_mysql_EDGE_input_for_edge_docker.tgz) and download it in the `EDGE-COVID19` folder.

```
$wget https://edge-dl.lanl.gov/EDGE/docker/edge_ubuntu_init_mysql_EDGE_input_for_edge_docker.tgz
```

Unzip the download file

```
$tar -xvzf edge_ubuntu_init_mysql_EDGE_input_for_edge_docker.tgz
```

Create Output and Report directories

```
$ mkdir -p EDGE_output  EDGE_report EDGE_input
```

Step 4: Start EDGE-COVID19 instance

Start the EDGE-COVID19 by running the following command in your Terminal from the `EDGE-COVID19` folder.

```
$ docker run -d -v $PWD/mysql:/var/lib/mysql \
 -v $PWD/EDGE_output:/home/edge/EDGE_output \
 -v $PWD/EDGE_input:/home/edge/EDGE_input \
 -v $PWD/EDGE_report:/home/edge/EDGE_report \
 -p 80:80 -p 8080:8080 --name edge_ncov bioedge/edge_ncov
```

Wait for a few minutes for the docker image to start the EDGE service; then Open http://localhost/ in a browser (Firefox, Chrome, Safari) to start the EDGE-COVID19. You will see the following screen.
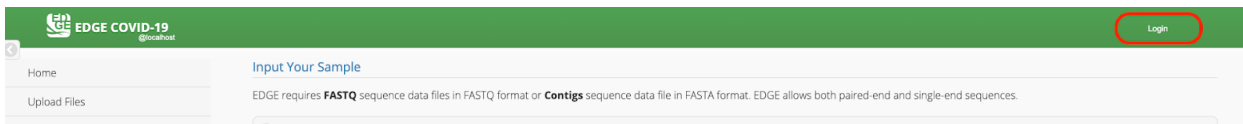

Step 5: Create a user account (if needed)

If EDGE-COVID19 is to be used by a single user then there is no need to create an account. You can log in directly using following credentials by clicking on the `Login` button on top right:

> **EDGE user**: admin_docker@my.edge
> **EDGE password**: admin_docker

Step 6: Upload your raw reads

The easiest way is to put it in your upload folder. The Upload folder can be found within the `EDGE_input` folder that you created in Step 3. Within `EDGE_input`, there will be a folder with a really long name and within that folder, there will be a folder called `MyUploads` where you can put your raw reads. This folder can then be seen from the web server (http://localhost/) by clicking on the button next to boxes where you input your FASTQ files.
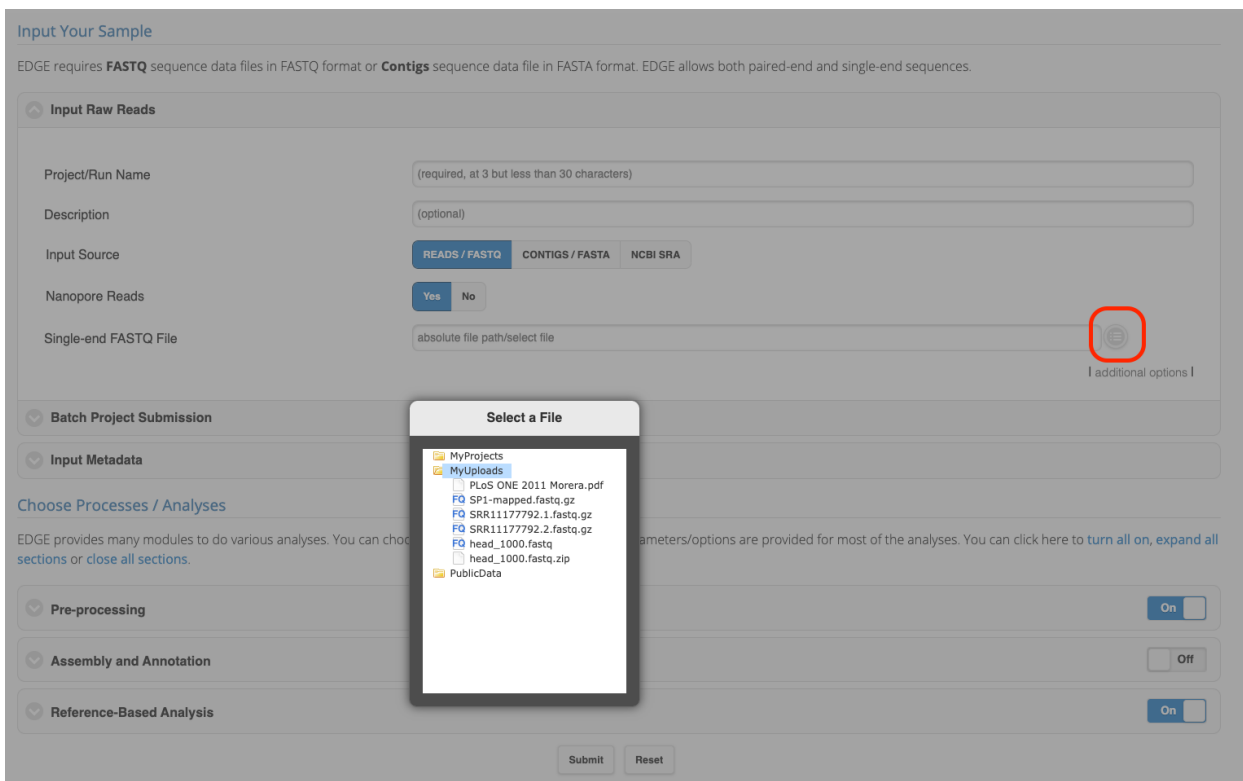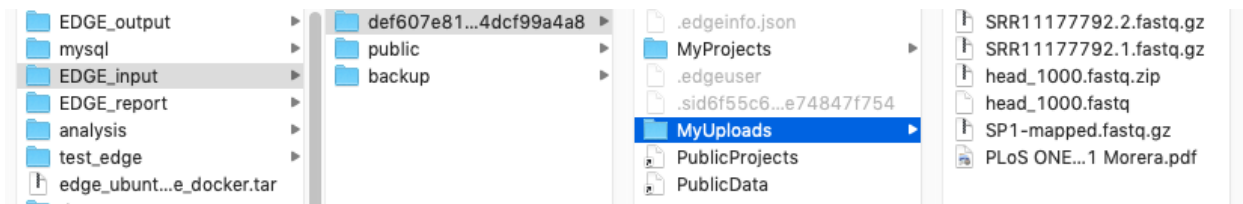
Step 7: Run your sample

If you are familiar with the EDGE Bioinformatics environment then you can skip this step and jump right into analyzing your data. Even if you are not familiar you should be still able to skip this step as EDGE Bioinformatics

has quite an intuitive design and you should be able to get on your analyses right away. However, here is a short description that will get you started. For detailed description you can also visit https://edge.readthedocs.io.

In the EDGE-COVID19 web server,
　　　　1. Type in a unique `Project/Run Name`.

　　　　2. Write in a short `Description`.
　　　　3. The nextIf you have your own raw reads select READS/FASTQ option.
　　　　4. If your sample was generated using Nanopore, select `Yes` in `Nanopore Reads, and if it's an illumina run, select `No`.
　　　　5. Input your raw reads by clicking on the button next to the input box and then navigate to your `MyUploads` folder where you have added your raw reads in Step 6.
　　　　6. You can skip the `Batch Project Submission`, if you are only doing one sample.
　　　　7. You can expand the `Input Metadata` section and fill in the metadata now so it's easier when you submit your genome.
　　　　8. If you want to do a QC on your reads, turn on the `Pre-processing` by sliding the button to the right. If you want to change other parameters you can expand the module by clicking on it and change parameters as you see fit.
　　　　9. If you prefer, you can also turn on `Assembly and Annotation`. Currently we have provided IDBA_UD, SPAdes, MEGAHIT, UniCycler, and LRASM.
　　　　10. `Reference-Based Analysis` should be turned ON by default. By default we use `bowtie2` aligner for illumina reads and `minimap2` aligners for Nanopore reads and do not perform variant analyses, and generate consensus sequence. You can change these parameters by expanding the module. `Variant Call` analyses can take quite a bit of time, especially for Nanopore reads. So, we recommend turning it off if it's really not needed.
　　　　11. Then, finally you can hit the submit button at the bottom of the page.

How to start/stop EDGE-COVID19?

　　　　To start the EDGE-COVID19 again, you will need to run following command in your Terminal from the same directory and then:

```
$ cd EDGE-COVID19
$ docker rm -v edge_ncov
$ docker run -d -v $PWD/mysql:/var/lib/mysql \
    -v $PWD/EDGE_output:/home/edge/EDGE_output \
    -v $PWD/EDGE_input:/home/edge/EDGE_input \
    -v $PWD/EDGE_report:/home/edge/EDGE_report \
    -p 80:80 -p 8080:8080 --name edge_ncov bioedge/edge_ncov
```

Wait a few minutes and go to http://localhost.

To stop the docker:

```
$ docker stop edge_ncov
```

How to update EDGE-COVID19?

To update the image to the latest version, you can pull the docker again in the original EDGE-COVID19 folder from **Step 1**.

```
$ docker pull bioedge/edge_ncov
```

After pulling the latest docker, we recommend restarting the edge_ncov container, and then starting the image from terminal:

```
$ cd EDGE-COVID19
$ docker rm -v edge_ncov
$ docker run -d -v $PWD/mysql:/var/lib/mysql \
    -v $PWD/EDGE_output:/home/edge/EDGE_output \
    -v $PWD/EDGE_input:/home/edge/EDGE_input \
    -v $PWD/EDGE_report:/home/edge/EDGE_report \
    -p 80:80 -p 8080:8080 --name edge_ncov bioedge/edge_ncov
```

How long will it take to run my sample?

We have compiled information on run times from different runs here:

[EDGE-COVID19 performance tracker](#)

What is in the output page?

An EDGE-COVID19

Contact:

You can view the discussions in the google group below and join the group to post questions or comments. EDGE user's google group at https://groups.google.com/d/forum/edge-users

Troubleshooting

Error in pulling image

If you have issues pulling this image, you may increase the basesize when launching docker daemon or use a different [Storage Driver] (https://goo.gl/8YUeUA). See [issue] (https://github.com/moby/moby/issues/8971).

IP address conflicts

Docker is hard coded to look for 172.17.0.1. If the ip address conflicts with the subnet of your wifi, you may need to customize the docker0 bridge by editing the /etc/docker/daemon.json as described [here] (https://forums.docker.com/t/change-default-docker0-bridge-ip-address/30470/9)

docker Commands Flags

* The -v /path/to/mysql:/var/lib/mysql part of the command mounts the /my/own/mysql (obtained from the link above and chown to mysql:mysql if needed) directory from the underlying host system as /var/lib/mysql inside the container, where MySQL by default will write its data files. Use this to persist the database data in the host. (You can use data volume instead and it is described in the "Note" section below.)
* The -v /path/to/EDGE_input:/home/edge/EDGE_input mounts the EDGE input directory structure (obtained from the git clone above) to persist the input/upload files/user projects in the host.
* The -v /path/to/EDGE_output:/home/edge/EDGE_output mounts the EDGE output directory to persist the output files in the host.
* The -v /path/to/EDGE_report:/home/edge/EDGE_report mounts the EDGE report directory to persist the report files in the host.
* The -p host:container binds the host ports 80 and 8080 to container ports 80 and 8080 inside the container. You can change the 80 and 8080 to fit your host system requirements.
* In some cases, you need to set the directory into 0777 mode. Please try opening up the directory permissions if you run into trouble.
* You can add "--restart always" in the docker run command to restart the container if it stops or when Docker daemon restarts.

* There is an issue mounting the host volume using MAC OSX. (https://github.com/boot2docker/boot2docker/issues/581). However, there is a workaround; use a data volume container instead.

```
    * To use a data volume container (a.k.a: OS independent.)
    $ docker pull bioedge/edge_ubuntu_mysql
    $ docker create --name mysql_data --volume /var/lib/mysql  bioedge/edge_ubuntu_mysql
    $ docker run -d --volumes-from mysql_data -v /path/to/EDGE_output:/home/edge/EDGE_output -v
/path/to/EDGE_input:/home/edge/EDGE_input/ -v /path/to/EDGE_report:/home/edge/EDGE_report -p 80:80 -p
8080:8080 --name edge_ncov bioedge/edge_ncov
```
Commands for checking status and error log

* Check container status.

```
  $ docker ps -a
```

* To Start/Stop container.
```
  $ docker start edge_ncov
  $ docker stop edge_ncov
```

* Check the mysql status in container:
```
    $ docker exec edge_ncov service mysql status
```
where "edge_ncov" is the container name when user `docker run` it with --name flag

* And user management system service status:
```
    $ docker exec edge_ncov service tomcat7 status
```
* For the Apache web server status and log:
```
    $ docker exec edge_ncov service apache2 status
    $ docker exec edge_ncov tail /var/log/apache2/error.log
    $ docker exec edge_ncov tail /var/log/apache2/access.log
```

Citation

Po-E Li, Chien-Chi Lo, Joseph J. Anderson, Karen W. Davenport, Kimberly A. Bishop-Lilly, Yan Xu, Sanaa Ahmed, Shihai Feng, Vishwesh P. Mokashi, Patrick S.G. Chain; Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform, Nucleic Acids Research, Volume 45, Issue 1, 9 January 2017, Pages 67–80, https://doi.org/10.1093/nar/gkw1027

Extraneous information

The user management can be accessed by http://localhost:8080/userManagement if the host port is 8080.

You can even disable the User management system which will then make it :

For security, you may need to update the credentials if the server will be used by others or by the public.