

EDGE COVID-19 documentation

Los Alamos National Laboratory

Email us: edge-covid19@lanl.gov

v1.0.12

Overview	3
A step by step guide to running EDGE COVID-19:	4
Step 1: Create an account	4
Step 2: Upload your raw reads	4
Step 3: Run your sample	6
EDGE COVID-19 output page	14
FAQs	17
For web based app:	17
How can I view alignments in a local viewer such as IGV?	17
For local docker build:	17
1. How to start/stop EDGE COVID-19 docker instance?	17
2. How to update EDGE COVID-19?	18
3. How long will it take to run my sample?	18
4. I am getting an error while pulling the image. What can I do?	18
6. IP address conflicts	19
7. What are the commands for checking status and error log?	19
8. How can I update the number of CPUs that EDGE COVID-19 uses?	19
Contact:	20
Citation:	20

Overview

EDGE COVID-19 is a tailored bioinformatics platform based on the more flexible and fully open-source [EDGE Bioinformatics](#) software ([Li et al. 2017](#)). This mini-version consists of a user-friendly GUI that drives standardized workflows for genome reference-based ‘assembly’ and preliminary analysis of Illumina or Oxford Nanopore (ONT) data for SARS-CoV-2 genome sequencing projects. **The result is a final SARS-CoV-2 genome ready for submission to [GISAID](#) and [GenBank](#).**

The default workflow in *EDGE COVID-19* includes:

- 1) data quality control (QC) and filtering,**
- 2) alignment of reads** to the original (first available) reference genome ([NC_045512.2](#), we removed the PolyA tail from the 3' end (33 nt)),
- 3) creation of a consensus genome sequence** based on the read alignments, and
- 4) a Single Nucleotide Polymorphism and Variant analyses**, with detail such as location and resulting coding differences.

The *EDGE COVID-19* platform can accommodate Illumina or ONT data, including ONT data from the [SARS-CoV-2 ARTIC network sequencing](#) protocols. Users can input/upload Illumina or Nanopore sequencing FASTQ files (and/or download from [NCBI SRA](#)). For Illumina data, default analyses include read QC, read mapping to the reference, and SNP/variant analysis. For ONT data, the data must be demultiplexed prior to uploading; the samples will be processed individually. The SNP/variant calling is not on by default for ONT. However, other functions (e.g. *de novo* assembly for whole genome data) are also available for both sequencing platforms. While command line execution is possible ([see here](#) and [here](#)), the GUI provides an easy data submission and results viewing platform, with the graphical and tabular views of variant/SNP data and a genome browser to view read coverage and location of SNPs or variants, as well as the reference annotations.

We have tested these workflows using Illumina (e.g. [SRR11393704](#)) and ONT (e.g. [SRR11397722](#)) datasets; these projects (along with a few others) are made public on the [site](#). This light-weight version is also available as a Docker container, able to run on any local hardware infrastructure.

Note: For *EDGE Bioinformatics* users who would also like to use the phylogeny or read- and assembly-based taxonomy classification tools to identify all organisms that may be present within complex samples, we recommend using the original [EDGE Bioinformatics](#) platform which harbors several tools and associated (large) databases that enable such a search. *In initial tests of taxonomy classification of SARS-CoV-2 samples (with no SARS-CoV-2 genomes in any of*

the databases), we recover SARS coronavirus and Bat coronavirus as the nearest neighbors (See table below).

Tool	#Reads	%Reads	Level	Top1	Top2	Top3	Top4	Top5	Columns...
gottcha-strDB-v	7,827	6.0	strain	SARS coronavirus	Bat coronavirus BM48-31/BGR/2008	N/A	N/A	N/A	
pangia	5,008	3.9	strain	SARS coronavirus	Bat coronavirus BM48-31/BGR/2008 strain BtCoV/BM48-31/BGR/2008	N/A	N/A	N/A	
metaphlan2	0	0.0	strain	N/A	N/A	N/A	N/A	N/A	
bwa	3,296	2.5	strain	Bat coronavirus Rp/Shaanxi2011	Bat coronavirus Cp/Yunnan2011	BtRs-BetaCoV/HuB2013	Rhinolophus affinis coronavirus	Bat SARS-like coronavirus RsSHC014	
kraken2	48,317	37.2	strain	SARS coronavirus	Rhodococcus opacus PD630	Burkholderia dolosa PC543	Xanthomonas citri pv. fuscans	Aeromonas hydrophila ML09-119	

A step by step guide to running *EDGE COVID-19*:

If you want to run the analyses in the cloud visit <https://edge-covid19.edgebioinformatics.org/> and follow the steps below:

Step 1: Create an account

You need to create an account. Click the “Sign up” link in the upper right corner of the page. After you have an account, you can click “Log in” for all subsequent visits and provide your user information.

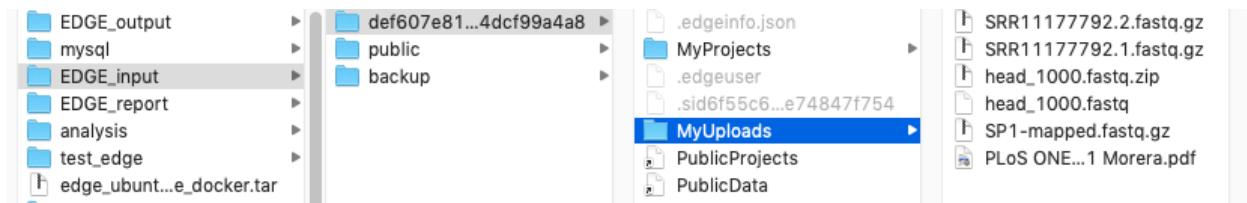


Step 2: Upload your raw reads

After you have logged in, you can click on “Upload Files” in the left menu. Drag and drop your data files into the window provided. Click “Start Upload” when you have added the files you need. The files will be put in a folder called MyUploads.

For a local installation:

The easiest way to upload your data is to put your data files in the upload folder, which can be found within the `EDGE_input` folder that you created when installing *EDGE COVID-19* docker [see [here](#)]. Within `EDGE_input`, there will be a folder with a long string of characters as the 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.



The `MyUploads` folder can be seen from the web server by clicking on the button to the right of the box(es) where you input your FASTQ files. (See figure below.) Click the file(s) you want to analyze.

Input Your Sample

EDGE requires **FASTQ** sequence data files in FASTQ format or **Contigs** sequence data file in FASTA format. EDGE allows both paired-end and single-end sequences.

Input Raw Reads

Project/Run Name	(required, at 3 but less than 30 characters)
Description	(optional)
Input Source	READS / FASTQ CONTIGS / FASTA NCBI SRA
Nanopore Reads	Yes No
Single-end FASTQ File	absolute file path/select file 

Batch Project Submission **Input Metadata**

Choose Processes / Analyses

EDGE provides many modules to do various analyses. You can choose sections or close all sections.

Pre-processing **Assembly and Annotation** **Reference-Based Analysis**

Select a File

MyProjects
MyUploads
PLoS ONE 2011 Morera.pdf
SP1-mapped.fastq.gz
SRR1117792.1.fastq.gz
SRR1117792.2.fastq.gz
head_1000.fasta
head_1000.fasta.zip
PublicData

Submit Reset

Step 3: 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 may still be able to skip this step, as EDGE Bioinformatics has a relatively intuitive design to instinctively get to your analyses right away. However, for completeness, here is a short description that will get you started. For a more detailed description, you can also visit our documentation site for EDGE Bioinformatics at <https://edge.readthedocs.io>.

In the *EDGE COVID-19* web server,

1. Type in a unique **Project/Run Name** with no spaces, but use underscores and/or dashes, if needed.
2. Write in a short **Description**. Spaces are allowed.
3. In the **Input Source** section, select **READS/FASTQ** for analyzing your own raw reads or **NCBI SRA** if you want to analyze COVID-19 samples deposited in SRA.
4. Select **Yes** in **Nanopore Reads** if your sample was generated using Nanopore; select **No** if it's Illumina data.
5. Input your raw reads by clicking on the button to the right of the input box (highlighted with a red box in the figure above) and then within the GUI navigate to your **MyUploads** folder where you have added your raw reads in Step 2.
6. You can skip the `Batch Project Submission`, if you are only processing one sample. A detailed instructions on using the batch mode can be found [here](#).

7. In the **Input Metadata** section, you can fill in the metadata so that you have all needed information when you are ready to submit genomes to NCBI or GISAID.
8. **Pre-processing** (Data QC) is turned **ON** by default and uses [FaQCs](#). This includes trimming low quality regions of reads and filtering reads that either fail a quality threshold or minimum length. If you wish to change parameters, you can expand the module by clicking on it and modify as desired. The default parameters are as follows:
 - Trim Quality Level: 20 (Illumina), 7(Nanopore)
 - Minimum Read Length: 50 (Illumina), 350(Nanopore)
 - "N" Base Cutoff: 10
 - Low Complexity Filter: 0.85
9. **Trimming primers from samples sequenced using multiplex amplicon approach.** We provide two options to trim primers if a multiplex amplicon approach such as [ARTIC \(v1-3\)](#), CDC protocols and [SWIFT protocols](#) were used. Default approach is to use the *align_trim*, that soft clips primer region from the alignment file (BAM) based on the position of primers in the reference genome. Another approach is to use FaQC, which trims the regions from reads that match with primer sequences.



10. For samples with whole genome sequencing (WGS) data that are interested in *de novo* assembly, you can also turn on `Assembly and Annotation`. Currently we provide [IDBA_UD v1.1.1](#), [SPAdes v3.13.0](#), [MEGAHIT v1.1.3](#), [UniCycler v0.4.8](#), [wtbdbg2 v2.5](#), [flye v2.8](#) and [miniasm v0.3](#) as options for assemblies.
11. **Reference-Based SARS-CoV-2 Genome Analysis** is turned **ON** by default. For Illumina data, [BWA mem](#) is used as the default aligner, which is then automatically followed by generation of a consensus sequence and variant calling. For ONT data, [minimap2](#) is the default aligner which is also automatically followed by generation of a consensus sequence, but not variant calling. We currently turn **OFF** variant calling for ONT data as it takes well over 24 hours on this platform. However, you can change any of these parameters by expanding the module and selecting the desired changes. Additionally, to avoid partial short alignment, [samclip](#) script with *max clip* length 50 for Illumina and 150 for ONT is also applied.

Variant calling: EDGE COVID-19 uses [bcftools mpileup](#) command to convert the aligned BAM file into genomic positions and call genotypes, reduce the list of sites to those found to be variants by passing this file into [bcftools call](#) command. The variant calls are filtered further by [vcfutils.pl](#) of SAMtools with following criteria:

- Minimum Root Mean Square (RMS) mapping quality for SNPs [10];

- minimum read depth [5];
- maximum read depth [300];
- minimum number of alternate bases [3];
- minimum ratio of alternate bases [0.3];
- SNP within INT bp around a gap to be filtered [3];
- ‘window size for filtering adjacent gaps [10];
- min P-value for end distance bias [0.0001];
- maximum fraction of reads supporting an indel [0.5];

The [consensus workflow](#): EDGE COVID-19 uses a maximum of 8000X depth coverage reads for computational efficiency. For samples sequenced using non-amplicon methods, PCR deduplication is also performed. Various parameters are defaulted including a minimum of 5x depth coverage of support or variant siter coverage per base (otherwise the consensus will be “N”), base quality (<20 for Illumina and <7 for ONT), alternate base Threshold (0.5 to support an alternative for the consensus to be changed), indels Threshold (to support an INDEL for the consensus to be changed, 0.5 for illumina and 0.6 for amplicon-based ONT), and minimum mapping quality of 60.

12. Click the **submit** button at the bottom of the page to start your job
13. The status of all of your projects can be viewed by clicking the **Projects** tab on the left menu and then clicking on **My Project List**. A detailed description can also be found [here](#).

The screenshot shows the 'Genome Analytics HOME' interface for the EDGE COVID-19 project. The top navigation bar includes the project logo and name (@edge-covid19.edgebioinformatics). Below the navigation bar, there is a sidebar with links: 'Genome Analytics HOME', 'About EDGE COVID-19', 'Upload Files', 'Run EDGE COVID-19', 'Get EDGE COVID-19', and 'Reports'. A 'Projects' tab is selected, indicated by a blue background and white text. Under the 'Projects' tab, there is a search bar labeled 'Find project by name/time' and a button labeled 'My Project List'. The main content area displays three project entries in a table:

Date	Project ID	Status
2020-04-23 01:48:41	SRR11494509_v3	✓ (Green)
2020-04-23 01:37:44	SRR11494509_v2	✓ (Green)
2020-04-22 22:06:30	SRR11494509_v1	✓ (Green)

If a project is highlighted in **green** it means the project has finished; if **orange**, **red**, or **grey** (not shown) it means the project is running, cancelled/failed, or not yet started, respectively. You can access the details and outputs of each project by clicking on the project in the list. Once selected, the project results page displays a summary of the run and general statistics of what tools and modules were activated and their runtime and status, as well as links to log files that provide detailed information on the command lines and parameters used for each tool executed. The rest of the page is divided by the modules that were originally selected for analysis. When selecting a project which has not finished running, some of the completed results may still be viewed on the page, however graphics and links to interactive features will not be present, as the rendering of figures is performed only in the last step.

14. Tree Placement of consensus genome(s) by UShER (Ultrafast Sample placement on Existing Tree): you can select the consensus genomes from the project list to do tree placement and output to a new tab with UShER result (link to UCSC).

Project Name	Status	Submission Time (MDT)	Total Running Time	Owner
2111_010	Complete	2020-10-09 11:24:22	00:20:53	Andrew Bartlow
2111_009	Complete	2020-10-09 11:24:07	00:17:07	Andrew Bartlow
2111_008	Complete	2020-10-09 11:23:52	00:35:11	Andrew Bartlow
2111_007	Complete	2020-10-09 11:23:37	00:25:04	Andrew Bartlow
2111_005	Complete			Andrew Bartlow
2111_006	Complete			Andrew Bartlow
2111_004	Complete			Andrew Bartlow
2111_003	Complete			Andrew Bartlow
2111_002	Complete			Andrew Bartlow
SRR12349131	Complete			Hajnalka Daligault
cats_2111_001_test	Complete			Andrew Bartlow
SRR11514749	Complete			Hajnalka Daligault
SRR11593395_no_primer_removal	Complete	2020-10-01 13:01:50	00:15:36	Hajnalka Daligault
SRR12110157	Complete	2020-10-01 11:19:03	00:18:47	Hajnalka Daligault

Fasta Sequence	Size (?)	#Ns (?)	#Mixed (?)	Bases aligned (?)	Insertions (?)	Deletions (?)	#SNVs used for placement (?)	#Masked SNVs (?)	Neighboring sample in tree (?)	Lineage of neighbor (?)	#Imputed values for mixed bases (?)	#Maximally parsimonious placements (?)	Parsimony score (?)	Subtree number (?)
SRR11514749_consensus_hCoV_19_SouthKorea_S2_2020_EPI_ISL_485393_2020_03_30_2020_07_09_South	29828 (?)	2461 (?)	0	27227 (?)	1268 (?)	1262 (?)	5 (?)	1 (?)	India/DL-NCDC-3982/2020 EPI_ISL_436445 20-04-09	B.6	0	18	0	1 (view in Nextstrain)
SRR11514749_consensus_NC_045512_2	29782 (?)	313 (?)	0	29469 (?)	0	0	6 (?)	1 (?)	India/DL-NCDC-3982/2020 EPI_ISL_436445 20-04-09	B.6	0	3	0	2 (view in Nextstrain)
SRR12110157_consensus_hCoV_19_USA_WA_UW_1631_2020_EPI_ISL_477693_2020_03_20_2020_06_29_USA	29864 (?)	2 (?)	0	29862 (?)	0	0	5 (?)	0	USA/CT_9849/2020 EPI_ISL_426416 20-03-08	A.1	0	1	0	3 (view in Nextstrain)
SRR12110157_consensus_NC_045512_2	29868 (?)	1 (?)	0	29867 (?)	0	0	5 (?)	0	USA/CT_9849/2020 EPI_ISL_426416 20-03-08	A.1	0	1	0	3 (view in Nextstrain)

Subtree 1: Unrelated sample

SRR11514749_consensus_hCoV_19_SouthKorea_S2_2020_EPI_ISL_485393_2020_03_30_2020_07_09_South

Differences from the reference genome (NC_045512.2): C6310A, C6312A, C13730T, C19524T, C23929T

Mutations along the path from the root of the phylogenetic tree to SRR11514749_consensus_hCoV_19_SouthKorea_S2_2020_EPI_ISL_485393_2020_03_30_2020_07_09_South:

C13730T > C26311T > C6312A > C23929T > C19524T > C6310A

This placement is not the only parsimony-optimal placement in the tree; 17 other placements exist.

Nearest neighboring GISAID sequence already in phylogenetic tree: India/DL-NCDC-3982/2020|EPI_ISL_436445|20-04-09: lineage B.6

15. Before submitting the pipeline run, you can prepare your genome for submission to GISAID and NCBI by inputting the metadata. You can do this after the pipeline is run as well.

EDGE COVID-19
@edge-covid19.edgebiobinformatics.org

Login / Sign up

- Upload Files
- Run EDGE COVID-19
- Get EDGE COVID-19
- Reports
- Projects

Input Raw Reads

Batch Project Submission

Input Metadata

Virus detail

Virus name: hCoV-19/Country/Identifier/2020

Passage details/history: Example: Original, Vero

Sample information

Collection date: Example: 2020-03-27, 2020-03 (collection in March, day unknown), 2020 (month and day unknown)

Location: Continent/Country/Region

Host: Example: Human, Environment, Canine, *Manis javanica*, *Rhinolophus affinis*, unknown

Gender: Example: Male, Female, or unknown

Patient age: Example: 65, 7 months, or unknown

After entering the metadata, the genome can be submitted to GISAID and NCBI directly through the *EDGE COVID-19* platform. You can access this functionality by clicking on the green check mark in the Reference-based results just below “Ready to Submit”.

Reference-Based SARS-CoV-2 Genome Analysis

a. Reads Mapped to Reference(s)

i. Mapped Reads By bwa

SARS-CoV2 Reference	Ref Length	Ref GC%	Mapped Reads	Mapped Reads%	Base Coverage	Avg Fold	Bam File
NC_045512.2	29,870	38.01%	119,708	99.04%	99.97%	725.92X	

Link to Directory

1 out of 1 reference(s) is(are) covered by input reads.

ii. Consensus Genome Statistics

SARS-CoV2 Reference	Consensus Length	Gaps	Ns/ns	5' Ns/ns	3' Ns/ns	SNVs	INDELs	Lineage	Consensus Genome	Ready to Submit
NC_045512.2	29,870	1	88	82	6	5	0	A.1		

Link to Directory

A menu will appear on the right side of the screen. Click the Metadata Action -> Upload to GISAID and NCBI option at the bottom of the menu to submit consensus genomes.

The screenshot shows the EDGE COVID-19 project interface. On the left, the project summary for SRR11241255 is displayed, including submission details and analysis status. The 'Job Progress' sidebar on the right lists completed steps with green checkmarks. A red arrow points to the 'Upload to GISAID and NCBI' option in the 'Metadata Action' menu.

Analysis	Run	Status	Running
Download SRA	On	Complete	00:00:10
Count Fastq	Auto	Complete	00:00:01
Quality Trim and Filter	On	Complete	00:00:04
Reads Mapping To Reference	On	Complete	00:00:30
Variant Analysis	Auto	Complete	00:00:01
Generate JBrowse Tracks	On	Complete	00:00:09
HTML Report	On	Complete	00:00:16

Report/Info	Location
Input Reads	SRR11241255
Output Directory	SRR11241255
PDF Report	final_report.pdf
MetaData	metadata.txt
Process log	process.log
Error log	error.log
Direct access	link

Job Progress

SRR11241255

- Download SRA
- Count Fastq
- Quality Trim and Filter
- Reads Mapping To Reference
- Variant Analysis
- Generate JBrowse Tracks
- HTML Report

Last checked: 2021-04-06 23:41:07

EDGE Server Usage

CPU	0.0 %
MEM	10.5 %
DISK	76.0 %

Action

- View live log
- Force this project to rerun
- Reconfig project (BETA)
- Interrupt running project
- Delete entire project
- Empty project outputs
- Share project
- Make project public
- Rename Project

Metadata Action

- Update Metadata
- Upload to GISAID and NCBI
- Upload to NCBI SRA

Submit SRR11241255_3 to GISAID

Virus detail

Virus name	hCoV-19/USA/LANL01/2020
Passage details/history	Original

Sample information

Collection date	2020-09-08
Location	North America/USA/Los Alamos
Host	Homo Sapiens
Gender	Male
Patient age	65
Patient status	Unknown
Sequencing technology	Illumina
Assembly method	EDGE-covid19: bwa 0.7.12-r1039, Consensus min coverage: 5X, min map quality: 60, Alternate Base > 50%, Indel > 50%.
Consensus Fasta	NC_048512.2 (99.9%, 728X), Ready to Submit

Institute information

Originating lab	New Mexico Department of Health Scientific Laboratory Division
Address	1101 Camino De Salud NE, Albuquerque, NM 87102
Submitting lab	LANL Bioscience
Address	PO 1663 MS888, Los Alamos, NM 87544
Authors	Migun Shakya, Chienchi Lo, Patrick Chan, Cheryl Gleasner, Alina Deshpande

Submitter information

Submitter	Chienchi Lo
GISAID ID	[]
GISAID Password	[]
NCBI ID	[]
NCBI Password	[]

Required metadata fields must be properly filled for submission to proceed.

Note: this feature is in beta format, and GISAID and NCBI can change the submission process at any time; if you run into any trouble, you can contact us at edge-covid19@lanl.gov.

16. Raw Reads submit to NCBI SRA:

In the same menu on the right side of the screen, users can submit the raw reads fastq to NCBI SRA. Click the Metadata Action -> Upload to NCBI SRA option at the bottom of the menu to submit.

The screenshot shows the project details for SRR11241255. The 'Job Progress' section lists various steps: Download SRA, Count Fastq, Quality Trim and Filter, Reads Mapping To Reference, Variant Analysis, Generate JBrowse Tracks, and HTML Report, all marked as complete with green checkmarks. The 'EDGE Server Usage' section shows CPU at 0.0%, MEM at 10.5%, and DISK at 76.0%. The 'Action' section includes options like View live log, Force this project to rerun, Reconfig project (BETA), Interrupt running project, Delete entire project, Empty project outputs, Share project, Make project public, and Rename Project. A red arrow points to the 'Upload to NCBI SRA' button in the 'Metadata Action' section, which is located at the bottom of the right sidebar.

Project Summary
Description: -
Submission Time: 2020 Sep 23 20:42:10 (MDT)
Number of CPUs: 10
Project Status: Complete
Total Analysis Run Time: 00:01:15
Last Run Time: 00:01:15
Owner: chienchi@lanl.gov

General

Analysis	Run	Status	Running
Download SRA	On	Complete	00:00:10
Count Fastq	Auto	Complete	00:00:01
Quality Trim and Filter	On	Complete	00:00:04
Reads Mapping To Reference	On	Complete	00:00:30
Variant Analysis	Auto	Complete	00:00:01
Generate JBrowse Tracks	On	Complete	00:00:09
HTML Report	On	Complete	00:00:16

Report/Info	Location
Input Reads	SRR11241255
Output Directory	SRR11241255
PDF Report	final_report.pdf
MetaData	metadata.txt
Process log	process.log
Error log	error.log
Direct access	link

SRA Sample Metadata

Action

- View live log
- Force this project to rerun
- Reconfig project (BETA)
- Interrupt running project
- Delete entire project
- Empty project outputs
- Share project
- Make project public
- Rename Project

Metadata Action

- Update Metadata
- Upload to GISAID and NCBI
- Upload to NCBI SRA

Submit SRR13361443 to NCBI SRA

The screenshot shows a form for submitting a BioProject to NCBI SRA. At the top, there's a section for "BioProject" with a "Use Registered BioProject" field containing "Yes" and "No" buttons. Below it is an "Existing BioProject" field with the value "PRJNA714680". Under "BioProject", there are sections for "BioSample", "Experiment", and "Additional Information", each with a dropdown arrow icon.

Required metadata fields must be properly filled for submission to proceed.

Note: this feature is in beta format, if you run into any trouble, you can contact us at edge-covid19@lanl.gov.

17. Batch submit (consensus genomes):

You can access this functionality by clicking on **My Project List**. Select on projects you would like to do the batch submission and then click on the right-most action button at top of the table.

The screenshot shows the "Project List" page of the EDGE COVID-19 platform. At the top, there's a header with the logo and the project ID "SRR11241255 / Chienchi". On the left is a sidebar with links to "Genome Analytics HOME", "About EDGE COVID-19", "Upload Files", "Run EDGE COVID-19", "Get EDGE COVID-19", "Reports", and "Projects". Below the sidebar is a search bar with "Find project by name/time" and a "My Project List" button. The main area is a table titled "Project List" with a "Submit Selected Projects to GISAID/NCBI" button above it. The table has columns for "Project Name", "Status", "Submission Time (MDT)", "Total Running Time", "Type", and "Owner". There are checkboxes next to each row. The rows for "ERR4868644" and "SRR13361443" have checkboxes checked and are highlighted in yellow. Other rows are greyed out. The bottom of the table shows the selected project names and their status: "ERR4868644 Complete" and "SRR13361443 Complete".

The action button will bring up the selected projects metadata table for users to fill in. (can scroll to the right to see other metadata.)

Selected Project Metadata

Type directly in the form below OR [upload a tab-delimited text file](#).

Show 25 entries

Project Name *	Virus Name	Passage Details	Collection Date	Location	Host	Gender	Patient Age	Patient Status	Sequencing Technology	Consensus Fasta	Submit
<input type="checkbox"/> ERR4868644	hCoV-19/USA/LANL01/2020	Original	2021-02-03	North America/USA	Homo S.	Female	56	Unknown	Illumina	NC_045512_2 (99.71%, 823X), Ready to Submit	<input type="button" value="Submit"/>
<input type="checkbox"/> SRR13361443	hCoV-19/USA/LANL02/2020	Original	2021-02-16	North America/USA	Homo S.	Other	56	Unknown	Illumina	NC_045512_2 (99.96%, 2458X), Ready to Submit	<input type="button" value="Submit"/>
<input type="checkbox"/> SRR13361443_o						Select Gender		Select Status		Select consensus genome	<input type="button" value="Submit"/>

Showing 1 to 3 of 3 entries

Don't see X scroll bar? On Mac, please try [this](#).

Additional information

Existing BioProject ?	[Optional] PRJNAXXXXX, ex:PRJNA714680
Release date	Example: 2021-04-20

Institute Information

Originating lab ?	New Mexico Department of Health Scientific Laboratory Division
Address	1101 Camino De Salud NE, Albuquerque, NM 87102
Submitting lab ?	Los Alamos National Laboratory Bioscience Division
Address	PO 1663 MS888, Los Alamos, NM 87545
Authors ?	Chien-Chi Lo, Migan Shakya, Cheryl Gleasner, Kim McMurry, Alina Deshpande, Twila Kunde, Joseph Hicks, Michael Edwards, Patrick Chain

Submitter information [?](#)

Submitter	Chienchi Lo
GISAID ID	chienchilo
GISAID Password	
NCBI ID	andy4748
NCBI Password	

Required metadata fields must be properly filled for submission to proceed.

Note: this feature is in beta format, and GISAID/NCBI can change the submission process at any time; if you run into any trouble, you can contact us at edge-covid19@lanl.gov.

18. Generate a report that contains comparison among multiple projects.

This feature can be accessed by clicking the “Reports” button on the left side of EC-19 page.

EDGE COVID-19 @edge-covid19.edgebioinformatics.org

SRR11241255 / Chiuchi

Genome Analytics HOME

About EDGE COVID-19

Upload Files

Run EDGE COVID-19

Get EDGE COVID-19

Reports

Report List

Create New Report

Projects

Find project by name/time

My Project List

New Report

Report Name (required)

Description (optional)

Choose Files/Figures/Stats for the Report

(please check all that apply):

Project/Run Information

Run Name Description Input Files Sample Metadata

Pre-processing

Table of stats Figures/Files

Assembly and Annotation

Table of stats Figures/Files

Reference-Based Analysis

Table of stats Figures/Files

Choose Projects for the Report

Show 5 entries Search:

<input type="checkbox"/>	Project Name	Status	Submission Time	Total Running Time	Type	Owner
<input type="checkbox"/>	ERR4868644	Complete	2021-02-17 16:42:11	00:02:07	private	Chiuchi Lo
<input type="checkbox"/>	ERR4893361	Complete	2021-02-12 22:19:18	00:03:37	private	Chiuchi Lo
<input type="checkbox"/>	SRR11241255	Complete	2021-02-03 23:23:07	00:00:50	shared	Chiuchi Lo
<input type="checkbox"/>	ERR4892461	Complete	2021-02-02 15:34:33	00:11:51	private	Chiuchi Lo
<input type="checkbox"/>	SRR13530301	Complete	2021-01-28 04:18:19	00:02:51	private	Chiuchi Lo
<input type="checkbox"/>	ERR4206007	Complete	2020-12-18 17:42:15			
<input type="checkbox"/>	SRR11445485	Complete	2020-12-18 15:50:40			
<input type="checkbox"/>			2020-11-20 21:13:11			

Showing 1 to 5 of 38 entries Previous 1 2 3 4 5 ... 8 Next

Reference-Based Analysis

1. Reads Mapped to Reference(s)

Mapped Reads

Link to [ref_reads.refs.csv](#)

Consensus Genome Statistics

Link to [ref_reads_cns.csv](#)

Variant Call

Link to [ref_reads_cns.csv](#)

Project/Run Name	Reference	Length	GC%	Mapped Reads	Mapped Reads%	Base Coverage	Avg Fold	Fold std.
SRR13361443_o	NC_045512.2	29,870	38.01%	532,889	99.71%	99.95%	2439.56X	1187.11X
SRR11547279	NC_045512.2	29,870	38.01%	12,804	23.92%	78.35%	142.14X	205.06X
SRR13361443	NC_045512.2	29,870	38.01%	506,204	94.72%	99.94%	2131.33X	1116.63X

Project/Run Name	Reference	Consensus Length	GC%	GAP	Ns/ns	5' Ns/ns	3' Ns/ns	SNPs	INDELS	LINEAGE	Ready to Submit
SRR13361443_o	NC_045512.2	29,851	37.95%	3	32	17	11	29	12	B.1.1.7	
SRR11547279	NC_045512.2	29,787	20.87%	148	12,919	1,616	202	33	60	B	
SRR13361443	NC_045512.2	29,851	37.91%	3	52	38	14	29	4	B.1.1.7	

EDGE COVID-19 output page

For a more detailed description of the EDGE bioinformatics output page, please refer to our full documentation [here](#). Each selected module will be displayed as a subsection, and detailed results may be found in each section. The Pre-processing section, for example, will have details on various statistics from all reads both before and after quality trimming and filtering. If assembly/annotation is selected, this module's output will include the assembled contigs as a Fasta file in addition to assembly metrics and annotation files.

In the **EDGE COVID-19** version of **Reference-based SARS-CoV-2 genome analysis**, an overview of the statistics and reference genome coverage is presented, including fold coverage (in graphical form along the length of the reference genome), as well as number of SNPs and gaps discovered (including those at the 5' and 3' ends of the reference genome). The [Pangolin Lineage](#) assignment also reported with hyperlink to [outbreak.info](#) for detailed information. A warning icon will be shown if it is a [Variant of Concern \(VOC\)](#) or [Variant of Interest \(VOI\)](#). If any

INDELs cause the frameshift in the CDS region, a warning icon will be shown too. You can directly download the consensus genome by clicking on the download icon. In our report, we also provide a quality check of the consensus genome by providing a green check mark if the resulting consensus genome is longer than 25kb, has coverage depth greater than 10X, and less than 5% of the genome is Ns. More data such as reference genome, BAM file, etc. can be accessed via the **Directory** link which allows access to all output files (e.g., there is an output file detailing the genomic location of SNPs or variant nucleotides, their prevalence within reads covering that position, any changes in translated amino acid composition, etc.). (See below)

a. Reads Mapped to Reference(s)

i. Mapped Reads By bwa

SARS-CoV2 Reference	Ref Length	Ref GC%	Mapped Reads	Mapped Reads%	Base Coverage	Avg Fold	Bam File
NC_045512.2	29,870	38.01%	532,900	99.71%	99.95%	2458.68X	

[Link to I Directory](#)

1 out of 1 reference(s) is(are) covered by input reads.

ii. Consensus Genome Statistics

SARS-CoV2 Reference	Consensus Length	Gaps	Ns/ns	5' Ns/ns	3' Ns/ns	SNVs	INDELs	Lineage	Consensus Genome	Ready to Submit
NC_045512.2	29,851	3	49	38	11	29	4	B.1.1.7		

[Link to I Directory](#)

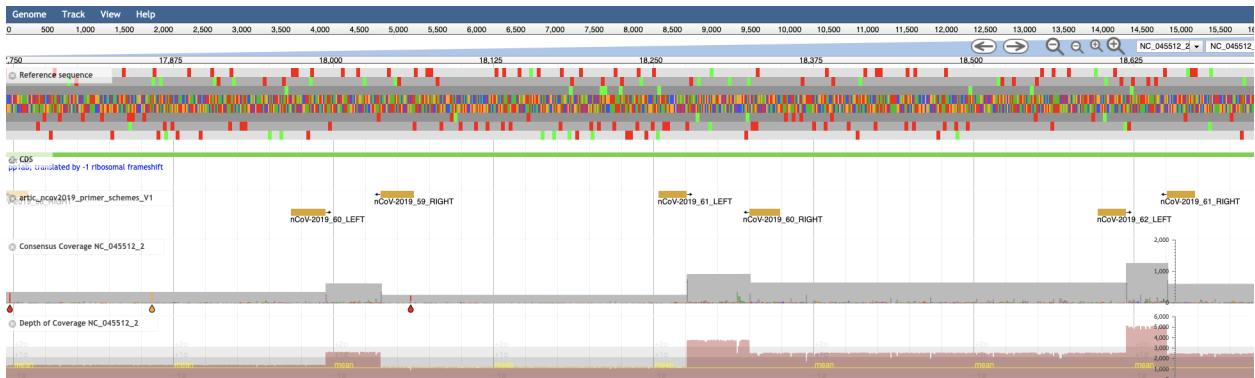
iii. Variant Call

SARS-CoV2 Reference	Variants	INDELs
NC_045512.2	29	1
NC_045512_2 Cov [full]		NC_045512_2 Fold [full]

Show the results in

[Link to I All Plots PDF](#) | [SNV Report](#) | [INDELs Report](#) | [Gap Table](#) | [Directory](#)

For scientists wishing to examine the details underlying the statistics, a JBrowse link is also provided (right below the graphics), which will open another browser window and allows interactive examination of the reference genome alignment results including annotations, locations of SNPs or variants, and read alignments (See below).



If the primer ***align_trim*** option has been used, users can also access the amplicon coverage plot in the output **Directory** and clicking the ***readsToRef_NC_045512_2_amplicon_coverage.html*** will open the graphics in a new browser window (See below).

a. Reads Mapped to Reference(s)

i. Mapped Reads

SARS-CoV2 Reference	Ref Length	Base Coverage	Avg Fold	Bam File
NC_045512.2	29,870	99.70%	1249.44X	

1 out of 1 reference(s) is(are) covered by input

ii. Consensus Genome Statistics

SARS-CoV2 Reference	Consensus Length	DELS	Consensus Genome	Ready to Submit
NC_045512.2	29,870			

Select a file

- NC_045512.2_consensus.fasta
- NC_045512.2_consensus.fasta.comp
- NC_045512.2_consensus.gaps
- NC_045512.2_consensus.gaps_report.txt
- consensus.log
- getConsensus.finished
- mapping.log
- readsToRef.alnstats.txt
- readsToRef.gaps
- readsToRef.stats.pdf
- readsToRef_NC_045512_2.coverage
- readsToRef_NC_045512_2.gap.coords
- readsToRef_NC_045512_2_amplicon_coverage.html** (selected)
- readsToRef_plots.pdf
- runReadsToGenome.finished
- variantAnalysis.finished
- variantAnalysis.log

[Link to Directory](#)

[Link to Directory](#)

NC_045512_2 Cov [full]

NC_045512_2 Cov [mean]



FAQs

For web based app:

1. How can I view alignments in a local viewer such as IGV?

You can download the BAM file using the green download button from the *Mapped Reads* section and the index file can be downloaded by clicking the hyperlinked **Directory** and then clicking on the index file (.bai).

2. Can edge-covid19 handle PacBio data?

Our QC tool FaQC cannot process the base quality values reported by PacBio Sequel as it reports all base qualities as PHRED 0. We recommend you run QC separately or turn **OFF** the **Preprocessing** module (FaQCs will throw an error due to Quality issue), select **YES** to **Nanopore Reads in Input Raw Reads** module, and add "-x map-pb" in the **Aligner Option** field in the additional options of **Reference-Based analysis module**.

For local docker build:

1. How to start/stop EDGE COVID-19 docker instance?

To start or restart *EDGE COVID-19*, you will need to run following command in your Terminal from the same directory:

```
$ cd EDGE-COVID19
$ docker rm -v edge-covid19
$ docker run -d --volumes-from mysql_data \
  -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-covid19 bioedge/edge-covid19
```

Then wait a few minutes and go to <http://localhost> in your favorite browser.

To stop the docker run following command in your directory:

```
$ docker stop edge-covid19
```

Note that the docker container will keep running in the background until you restart your computer or specifically stop it using the above command.

2. How to update EDGE COVID-19?

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

```
$ docker pull bioedge/edge-covid19
```

After pulling the latest docker, start the image from terminal:

```
$ cd EDGE-COVID19
$ docker rm -v edge-covid19
$ docker run -d --volumes-from mysql_data \
  -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-covid19 bioedge/edge-covid19
```

3. How long will it take to run my sample?

Using a Macbook Pro with 16GB RAM and 8 processors available:

dataset size (bases)	# Raw reads	Type of data (Nanopore/ Illumina)	Protocol	# of CPUs	Total Wall Clock Time
416,793,360	10,493,168	Illumina	Amplicons	4	2:38:01
594,064,863	1,382,016	Nanopore	ARTIC protocol	4	0:21:07

4. I am getting an error while pulling the image. What can I do?

If you have issues pulling this image, you may increase the basesize when launching docker daemon or use a different [Storage Driver](#). See similar [issue](#) here.

5. I am getting an error while trying to login on GUI: “Failed to login in. Please check server log for details”

In some linux environments, users need to set the /path/to/mysql directory into 0777 mode.

Please try opening the directory permissions if you run into trouble.

```
$ 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 $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-covid19 bioedge/edge-covid19
```

6. 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 docker bridge by editing the /etc/docker/daemon.json as described [here](#).

7. What are the commands for checking status and error log?

Check the MySQL status in container:

```
$ docker exec edge-covid19 service mysql status
```

where "edge-covid19" is the container name when using `docker run` with `--name` flag

Check container status.

```
$ docker ps -a
```

Check user management system service status:

```
$ docker exec edge-covid19 service tomcat7 status
```

Check the Apache web server status and log:

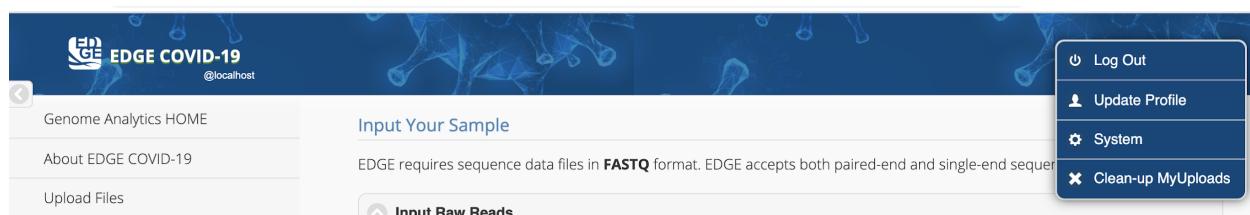
```
$ docker exec edge-covid19 service apache2 status  
$ docker exec edge-covid19 tail /var/log/apache2/error.log  
$ docker exec edge-covid19 tail /var/log/apache2/access.log
```

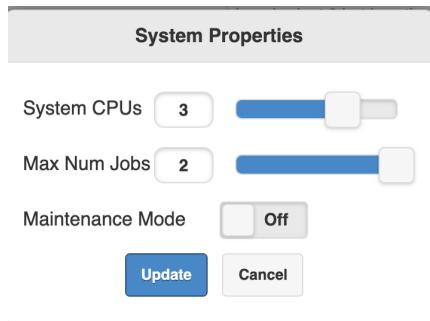
8. How can I update the number of CPUs that EDGE COVID-19 uses?

*

The default number of CPUs available to EDGE inside the container is 4 and the maximum number of jobs can run simultaneously is 2.

Each job will use $(\text{edge_system_cpu-1})/\text{max_num_jobs}$ in integer CPUs (as you see on the GUI). These numbers can be changed by login using an admin account and click on the user name to pop up the user menu where you can click the **System** button to open the system properties menu.





Contact:

You can view the discussions in the google group below and join the group to post questions and/or comments.

EDGE user's google group at <https://groups.google.com/d/forum/edge-users>

You can also directly contact us through email at edge-covid19@lanl.gov

Citation:

Lo, Chien-Chi, Migun Shakya, Karen Davenport, Mark C. Flynn, Jason D. Gans, Adán Myers y Gutiérrez, Bin Hu, Po-E Li, Elais Player Jackson, Yan Xu and Patrick S. G. Chain. "EDGE COVID-19: A Web Platform to generate submission-ready genomes for SARS-CoV-2 sequencing efforts." (2020).

<https://arxiv.org/abs/2006.08058>